

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 15:42:09 ; Search time 26.4966 seconds
(without alignments)
750.748 Million cell updates/sec

Title: US-09-993-059-18

Perfect score: 2322

Sequence: 1 MOLRNPGLHLCALALRELA.....RSHINPTCTVLIQSEKDEL 423

Scoring table: BLOSUM62

Gapop' 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2293	98.8	429	1 AGAL_HUMAN	P06280 homo sapien
2	1806	77.8	419	1 AGAL_MOUSE	P51569 mus musculus
3	952.5	41.0	411	1 NAGA_HUMAN	P17050 homo sapien
4	712	30.7	378	1 AGAL_COFAR	Q42656 coffea arab
5	681.5	29.3	411	1 AGAL_CYPATE	P14749 cyamopsis t
6	543.5	23.4	545	1 AGAL_ASPNG	P28351 aspergillus
7	521	22.4	471	1 MEL6_YEAST	P41947 saccharomyc
8	519	22.4	471	1 MEL5_YEAST	P41946 saccharomyc
9	510	22.0	471	1 MEL2_YEAST	P41945 saccharomyc
10	502	21.6	471	1 MEL1_YEAST	P04824 saccharomyc
11	129	5.6	641	1 IMD_ARTGO	Q44052 arthrobacte
12	125	5.4	348	1 Y113_AQUAE	O66516 aquifex aeo
13	114.5	4.9	438	1 AM3B_ORYSA	P27937 oryza sativ
14	114	4.9	733	1 AGAL_PEDPE	P43467 pediococcus
15	100.5	4.3	708	1 RAFA_ECOLI	P16551 escherichia
16	99.5	4.3	872	1 GUXA_CELFI	P50401 cellulomona
17	97.5	4.2	437	1 AM3C_ORYSA	P27939 oryza sativ
18	94.5	4.1	533	1 ARSB_HUMAN	P15848 homo sapien
19	93	4.0	592	1 VATA_SULSO	Q9UHW6 sulfolobus
20	91	3.9	622	1 YAK8_SCHPO	Q09920 schizosacch
21	91	3.9	966	1 ENV_CAEVC	P31626 caprine art
22	90	3.9	1271	1 RBL2_ARATH	Q9MAG6 arabidopsis
23	89.5	3.9	719	1 AGA2_PEDPE	P43469 pediococcus
24	89.5	3.9	1443	1 DPO3_MYCPN	P75080 mycoplasma
25	88.5	3.8	301	1 LEC1_MOUSE	P24721 mus musculus
26	88.5	3.8	461	1 ARU1_BACSU	O34858 bacillus su
27	88.5	3.8	1984	1 YL1_DROME	P98163 drosophila
28	88	3.8	855	1 RDL5_ARATH	Q8W3J8 arabidopsis
29	87.5	3.8	1246	1 NARG_ECOLI	P09152 escherichia
30	87.5	3.8	4349	1 DVHC_FUSSO	P78716 fusarium so
31	87	3.7	438	1 AMY1_HORVU	P00693 hordeum vul
32	87	3.7	758	1 PMT2_YEAST	P31382 saccharomyc
33	87	3.7	908	1 RPP8_ARATH	Q8W4J9 arabidopsis

RESULT 1

AGAL_HUMAN	STANDARD;	PRT;	429 AA.
ID	AC	P06280;	
DT	01-JAN-1988	(Rel. 06, Created)	
DT	01-AUG-1988	(Rel. 08, Last sequence update)	
DT	15-SEP-2003	(Rel. 42, Last annotation update)	
DE	Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase alfa).		
DE	GIA.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]	SEQUENCE FROM N.A.	
RP	TISSUE=Lymphoblast;		
RC	MEDLINE=89263745; PubMed=2542896;		
RX	Kornreich R., Desnick R.J., Bishop D.F.;		
RA	"Nucleotide sequence of the human alpha-galactosidase A gene.";		
RT	Nucleic Acids Res. 17:3301-3302(1989).		
RL	[2]	SEQUENCE FROM N.A.	
RN	TISSUE=Fibroblast;		
RP	MEDLINE=87246603; PubMed=3036505;		
RC	Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,		
RX	Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,		
RA	Ginns E.I.;		
RT	"Signal sequence and DNA-mediated expression of human lysosomal		
RT	alpha-galactosidase A.";		
RL	Eur. J. Biochem. 165:275-280(1987).		
RL	[3]	SEQUENCE FROM N.A.	
RP	MEDLINE=95352959; PubMed=7626884;		
RX	Oeltyen J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,		
RA	Gibbs R.A.;		
RT	"Sixty-nine kilobases of contiguous human genomic sequence containing		
RT	the alpha-galactosidase A and Bruton's tyrosine kinase loci.";		
RL	Mamm. Genome 6:334-338(1995).		
RN	[4]	SEQUENCE FROM N.A.	
RP	Willson S.;		
RA	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
RL	[5]	SEQUENCE FROM N.A.	
RP	TISSUE=uterus;		
RC	MEDLINE=22388257; PubMed=12477932;		
RX	Stausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		

ALIGNMENTS

34	86.5	3.7	535	1 ARSB_FELCA	P33727 felis silve
35	86	3.7	595	1 VATA_SULTO	Q971b7 sulfolobus
36	86	3.7	642	1 VP4_AHSV4	Q64929 african hor
37	84.5	3.6	361	1 RE1_HAEIN	P43917 haemophilus
38	84.5	3.6	535	1 DEXE_STRPN	Q54796 streptococc
39	84.5	3.6	660	1 TP6B_METJA	Q58434 methanococc
40	84	3.6	591	1 VATA_SULAC	P09639 sulfolobus
41	83.5	3.6	445	1 LAMB_AERSA	Q44287 aeromonas s
42	83.5	3.6	1049	1 RDL6_ARATH	Q94hw3 arabidopsis
43	83	3.6	507	1 AHPE_STAAT	O05204 staphylococ
44	83	3.6	678	1 RNB_VIBCH	Q9k1el vibrio chol
45	82.5	3.6	259	1 Y305_CHLMU	Q9p103 chlamydia m

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny K.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL hum. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RN SEQUENCE OF 31-429 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Desnick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
RN [7]
RN SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE=88112869; PubMed=2892762;
RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
RT human lysosomal enzyme, alpha-galactosidase A.";
RL Gene 58:177-188(1987).
RN [8]
RN SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE=98234528; PubMed=2836863;
RA Bishop D.F., Kornreich R., Desnick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
RT further evidence for the absence of a 3' untranslated region.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907(1988).
RN [9]
RN RNA EDITING
RX MEDLINE=95380278; PubMed=7503918;
RA Novo F.J., Kruszewski A., Mcdermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine
RT to purine conversion.";
RL Nucleic Acids Res. 23:2636-2640(1995).
RN [10]
RN REVIEW ON FD VARIANTS.
RX MEDLINE=94258158; PubMed=7911050;
RA Eng C.M., Desnick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
RT human alpha-galactosidase A gene.";
RL Hum. Mutat. 3:103-111(1994).
RN [11]
RN VARIANT FD SER-40.
RX MEDLINE=90092580; PubMed=2152885;
RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
RA Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
RT activity caused by a single amino acid substitution of Pro-40 by
RT Ser.";
RL FEBS Lett. 259:353-356(1990).
RN [12]
RN VARIANT FD VAL-296.
RX MEDLINE=91101674; PubMed=1846223;
RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
RA Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
RT to the myocardium.";
RL New Engl. J. Med. 324:395-399(1991).
RN [13]
RN VARIANT FD GLN-301.
RX MEDLINE=91022721; PubMed=2171331;
RA Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
RA Bishop D.F., Desnick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
RT in classical and atypical hemizygotes with Fabry disease.";
RN [14]

Am. J. Hum. Genet. 47:784-789(1990).
RN [14]
RN VARIANT FD TRP-356.
RX MEDLINE=89198098; PubMed=2539398;
RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
RA Sakuraba H., Desnick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
RT in the alpha-galactosidase gene.";
RL J. Clin. Invest. 83:1390-1399(1989).
RN [15]
RN VARIANTS FD GLN-66; CYS-112; GLU-279; GLN-301 AND ARG-328.
RX MEDLINE=92250012; PubMed=1315715;
RA Ishii S., Sakuraba H., Suzuki Y.;
RT "Point mutations in the upstream region of the alpha-galactosidase A
RT gene exon 6 in an atypical variant of Fabry disease.";
RL Hum. Genet. 89:29-32(1992).
RN [16]
RN VARIANTS SER-34; GLY-56; ARG-162; GLN-227; VAL-264; PHE-297;
RX TYR-313; ALA-328 AND ARG-404 DEL.
RX MEDLINE=94070864; PubMed=7504405;
RA Eng C.M., Resnick-Silverman L.A., Niehaus D.J., Astrin K.H.,
RA Desnick R.J.;
RT "Nature and frequency of mutations in the alpha-galactosidase A gene
RT that cause Fabry disease.";
RL Am. J. Hum. Genet. 53:1186-1197(1993).
RN [17]
RN VARIANTS FD SER-34; SER-215; ALA-269; LYS-327 AND ARG-361.
RX MEDLINE=93372810; PubMed=8395937;
RA Davies J.P., Winchester B.G., Malcolm S.;
RT "Mutation analysis in patients with the typical form of
RT Anderson-Fabry disease.";
RL Hum. Mol. Genet. 2:1051-1053(1993).
RN [18]
RN VARIANTS FD ARG-35; LEU-49; VAL-165 AND GLU-316.
RX MEDLINE=94348524; PubMed=8069316;
RA Davies J.P., Christomanou H., Winchester B.G., Malcolm S.;
RT "Detection of 8 new mutations in the alpha-galactosidase A gene in
RT Fabry disease.";
RL Hum. Mol. Genet. 3:667-669(1994).
RN [19]
RN VARIANTS FD
RX MEDLINE=95152501; PubMed=7531540;
RA Eng C.M., Niehaus D.J., Enriquez A.L., BURGERT T.S., Ludman M.D.,
RA Desnick R.J.;
RT "Fabry disease: twenty-three mutations including sense and antisense
RT Cpg alterations and identification of a deletion hot-spot in the
RT alpha-galactosidase A gene.";
RL Hum. Mol. Genet. 3:1795-1799(1994).
RN [20]
RN VARIANTS FD GLN-66; CYS-112; VAL-156; VAL-166; ALA-260; GLU-279;
RX ILE-296; GLN-301; LYS-320; ARG-328 AND SER-373.
RX MEDLINE=96024628; PubMed=7575533;
RA Okumura T., Ishii S., Takenaka T., Kase R., Kamei S., Sakuraba H.,
RA Suzuki Y.;
RT "Galactose stabilizes various missense mutants of alpha-galactosidase
RT in Fabry disease.";
RL Biochem. Biophys. Res. Commun. 214:1219-1224(1995).
RN [21]
RN VARIANTS FD TYR-142; VAL-156 AND VAL-166.
RX MEDLINE=95278912; PubMed=7759078;
RA Okumura T., Ishii S., Kase R., Kamei S., Sakuraba H., Suzuki Y.;
RT "Alpha-galactosidase gene mutations in Fabry disease: heterogeneous
RT expressions of mutant enzyme proteins.";
RL Hum. Genet. 95:557-561(1995).
RN [22]
RN VARIANTS FD PRO-32; SER-34; ASP-85; THR-156 AND GLN-301.
RX MEDLINE=95322989; PubMed=7599642;
RA Madsen K.M., Hasholt L., Soerensen S.A., Lagerstroem Fermer M.,
RA Dahl N.;
RT "Two novel mutations (L32P) and (G85N) among five different missense
RT mutations in six Danish families with Fabry's disease.";
RL Hum. Mutat. 5:277-278(1995).
RN [23]

```
Query Match          98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 9.7e-193;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEP 60
DB 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEP 60
OY 61 DSCISEKLFMEAEMLVSEGWKRDAGYEYLCIDCCWMAFQDSEGRGLQADPQRFPHGIROL 120
DB 61 DSCISEKLFMEAEMLVSEGWKRDAGYEYLCIDCCWMAFQDSEGRGLQADPQRFPHGIROL 120
OY 121 ANYVHSGKLGIGIYADVGNKTCAGPPGSGFYDYIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPPGSGFYDYIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
OY 161 ADGKYHMSLALNRTGRSIVYSCWPLYMWPFOKPNYETIROYCNHWRNFADIDDSWKSJK 240
DB 161 ADGKYHMSLALNRTGRSIVYSCWPLYMWPFOKPNYETIROYCNHWRNFADIDDSWKSJK 240
OY 181 SILDWTSFNQRIYDVAGPGGWNDDMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQRIYDVAGPGGWNDDMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
OY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGNDFEVRPLSGLAWAVAMINRQETG 360
DB 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGNDFEVRPLSGLAWAVAMINRQETG 360
OY 361 GPRSVTIYASVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSVTIYASVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 2
AGAL_MOUSE
ID AGAL_MOUSE STANDARD; PRT; 419 AA.
AC P51569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE Alpha-galactosidase) (Alpha-D-galactosidase A).
GN GLA OR AGS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=96125203; PubMed=8543175;
RA Ohshima T., Murray G.J., Nagle J.W., Quirk J.M., Kraus M.H.,
RA Barton N.W., Brady R.O., Kulkarni A.B.;
RT "Structural organization and expression of the mouse gene encoding
RT alpha-galactosidase A.";
RL Gene 166:277-280(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C129;
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci.";
RL Mamm. Genome 6:334-338(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96316016; PubMed=8733892;
RA Gotlib R.W., Bishop D.F., Wang A.M., Zeldner K.M., Ioannou Y.I.,
RA Adler D.A., Distche C.M., Desnick R.J.;
RT "The entire genomic sequence and cDNA expression of mouse alpha-
RT galactosidase A.";
```

```
RL Biochem. Mol. Med. 57:139-148(1996).
CC -!- CATALYTIC ACTIVITY: Melibiose + H(2)O = galactose + glucose.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U34071; AAA96749.1; -
CC EMBL; L46651; AAA74453.1; -
CC EMBL; U58105; AAB47244.1; -
CC EMBL; U50716; AAC52584.1; -
CC EMBL; U50715; AAC52583.1; -
CC PIR; JC4522; JC4522.
CC MGD; MGI:1347344; Glia.
CC InterPro; IPR002241; Glyco_hydro_27.
CC InterPro; IPR001111; Glyco_hydro_GHD.
CC Pfam; PF02065; Melibiase; 1.
CC PRINTS; PR00740; GLHYDRLASE27.
CC ProDom; PD002572; Glyco_hydro_GHD; 1.
CC PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.
FT SIGNAL 1 31 BY SIMILARITY.
FT CHAIN 32 419 ALPHA-GALACTOSIDASE A.
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 419 AA; 47642 MW; BD5E6A99AC113613 CRC64;

Query Match          77.8%; Score 1806; DB 1; Length 419;
Best Local Similarity 78.2%; Pred. No. 3.4e-150;
Matches 326; Conservative 41; Mismatches 50; Indels 0; Gaps 0;

OY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEP 60
DB 1 MKLLSRDRLVCELALCPALVFWLSILGVRLDNGLARTPTMGWLHWFEMCNLDQCEP 60
OY 61 DSCISEKLFMEAEMLVSEGWKRDAGYEYLCIDCCWMAFQDSEGRGLQADPQRFPHGIROL 120
DB 61 DACISEQLFMQMAELVSDGWRDAGYDYLCIDCCWMAFQDSEGRGLQADPQRFPHGIROL 120
OY 121 ANYVHSGKLGIGIYADVGNKTCAGPPGSGFYDYIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPPGSGFYDYIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
OY 181 ADGKYHMSLALNRTGRSIVYSCWPLYMWPFOKPNYETIROYCNHWRNFADIDDSWKSJK 240
DB 181 ENGKYMALALNRTGRSIVYSCWPLYMWPFOKPNYETIROYCNHWRNFADIDDSWKSJK 240
OY 241 SILDWTSFNQRIYDVAGPGGWNDDMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
DB 241 NTLSTVTVVQKEIVEVAGPGGWNDDMLVIGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
OY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGNDFEVRPLSGLAWAVAMINRQETG 360
DB 301 RQISSOKALLQDKDVIAINQDPLGKQGYCYCFKRNHIEVWRPLSNLAWAVARNLOETG 360
OY 361 GPRSVTIYASVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPCPTTIQISSLGRGLACNPGCIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 3
NAGA_HUMAN
ID NAGA_HUMAN STANDARD; PRT; 411 AA.
AC P17050;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-N-acetylgalactosaminidase precursor (EC 3.2.1.49) (Alpha-
 DE galactosidase B).
 GN NAGA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=91072392; PubMed=2174888;
 RA Wang A.M., Bishop D.F., Desnick R.J.;
 RT "Human alpha-N-acetylgalactosaminidase-molecular cloning, nucleotide
 RT sequence, and expression of a full-length cDNA. Homology with human
 RT alpha-galactosidase A suggests evolution from a common ancestral
 RT gene";
 RL J. Biol. Chem. 265:21859-21866(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=91257820; PubMed=1646157;
 RA Wang A.M., Desnick R.J.;
 RT "Structural organization and complete sequence of the human alpha-N-
 RT acetylgalactosaminidase gene: homology with the alpha-galactosidase A
 RT gene provides evidence for evolution from a common ancestral gene";
 RL Genomics 10:133-142(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=89392067; PubMed=2551294;
 RA Tsuji S., Yamauchi T., Hiraiwa M., Isobe T., Okuyama T., Sakimura K.,
 RA Takahashi Y., Nishizawa M., Uda Y., Miyatake T.;
 RT "Molecular cloning of a full-length cDNA for human alpha-N-
 RT acetylgalactosaminidase (alpha-galactosidase B)";
 RL Biochem. Biophys. Res. Commun. 163:1498-1504(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90321233; PubMed=2372288;
 RA Yamauchi T., Hiraiwa M., Kobayashi H., Uda Y., Miyatake T., Tsuji S.;
 RT "Molecular cloning of two species of cDNAs for human alpha-N-
 RT acetylgalactosaminidase and expression in mammalian cells";
 RL Biochem. Biophys. Res. Commun. 170:231-237(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baggeley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,

RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.-C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22";
 RL Nature 402:489-495(1999).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP VARIANT LYS-325
 RX MEDLINE=91056183; PubMed=2243144;
 RA Wang A.M., Schindler D., Desnick R.J.;
 RT "Schindler disease: the molecular lesion in the alpha-N-
 RT acetylgalactosaminidase gene that causes an infantile neuroaxonal
 RT dystrophy";
 RL J. Clin. Invest. 86:1752-1756(1990).
 RN [8]
 RP VARIANT CYS-160
 RX MEDLINE=96375755; PubMed=8782044;
 RA Keulemans J.L.M., Reuser A.J.J., Kroos M.A., Willemsen R.,
 RA Hermans M.M.P., van den Ouweland A.M.W., de Jong J.G.N., Wevers R.A.,
 RA Renter W.O., Schindler D., Coll M.J., Chabas A., Sakuraba H.,
 RA Suzuki Y., van Diggelen O.P.;
 RT "Human alpha-N-acetylgalactosaminidase (alpha-NAGA) deficiency: new
 RT mutations and the paradox between genotype and phenotype";
 RL J. Med. Genet. 33:458-464(1996).
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
 CC acetyl-D-galactosamine residues in N-acetyl-alpha-D-
 CC galactosaminides.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- DISEASE: Defects in NAGA are the cause of Schindler disease
 CC [MIM:104170], an autosomal recessive disorder characterized by the
 CC increased urinary excretion of glycopeptides and oligosaccharides
 CC containing alpha-N-acetylgalactosamine moieties. Histologically
 CC characterized by neuroaxonal dystrophy and clinically by
 CC neurological signs (convulsion during fever, epilepsy, psychomotor
 CC retardation and hypotonia).
 CC -1- DISEASE: Defects in NAGA are the cause of Kanizaki disease
 CC [MIM:104170]; also known as Schindler disease type II. It is an
 CC autosomal recessive disorder very similar to Schindler disease.
 CC Kanizaki disease is histologically characterized by vacuolization,
 CC and clinically by a late onset, intellectual impairment,

Db 16 LANGLGLTPMGWNSWNHRCNLD-----EKLIRETADAMVSKGLAALGYKYINL 65
 QY 92 DDCWMAQPDSEGRLOADPQREPHGIRQLANYVHSGKLGKIYADVGNKTKCA-GFPGSFG 150
 Db 66 DDCWAEINROSGNLVPGKSTFGGKALADYVHSGKLGKIYSDAGTQCTSKTMPGSLG 125
 QY 151 YYDIDAQTFADGWGVDLLKFDGCGYCDLSLADGYKHMSLALNRTGRSIVYS-CWPLYMW 209
 Db 126 HEQDAKTFASGWGVDLYKYDNCNNNI-SPKERYPIMSKALLNSGRSIFSLCEWG---- 180
 QY 210 PFQKPNYTEIQVCNHNWRNADIDDSWKSISILD-----WTSNQERIVDVAGPGWNDP 265
 Db 181 --BEDPATWAKEVGNWSRTTGDIDDSWSSMTSRADMDNKWASY-----AGPGWGNPD 230
 QY 266 DMLVIGNFGLSNQOVTOMALWMAAPLWMSNDLRHISPOAKKALODKDVIAINQDPLG 325
 Db 231 DMLVEVNGGMMTTEYRSHFSIWALAKAPLLIGCDIRSMGATFOLLSNAEVIADVQDKLG 290
 QY 326 KQGYQLRQGNFVEWPERPLSGLAWAVAMINR-----OEIGGPRSYTIAVASL 372
 Db 291 VQGNKVKTYGDLVWAGPLSKRVAVALWNRGSGSTATITAYWSDVGLPSTAVVYNARDL 348

RESULT 5

AGAL_CYATE STANDARD; PRT; 411 AA.
 AC P14749;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Alpha-galactosidase precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-galactoside galactohydrolase).
 OS Cyamopsis tetragonoloba (Guar) (Cluster bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Indigoferae;
 CC Cyamopsis.
 OX NCBI_TaxID=3832;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Aleurone;
 RX MEDLINE=91370836; PubMed=2577496;
 RA Overbeek N., Fellinger A.J., Toonen M.Y., van Wassenaar D.,
 RA Verrips C.T.;
 RT "Cloning and nucleotide sequence of the alpha-galactosidase cDNA from
 RL Cyamopsis tetragonoloba (guar).";
 RL plant Mol. Biol. 13:541-550(1989).
 RN [2]
 RP SEQUENCE OF 48-57 AND 172-178.
 RC TISSUE=Seed;
 RA Hughes S.G., Overbeek N., Robinson S., Pollock K., Smeets F.L.M.;
 RA "Messenger RNA from isolated aleurone cells directs the synthesis of
 RT an alpha-galactosidase found in the endosperm during germination of
 RT guar (Cyamopsis tetragonoloba) seed.";
 RL Plant Mol. Biol. 11:783-789(1988).
 CC -!- FUNCTION: INVOLVED IN THE HYDROLYSIS OF THE GALACTOMANNAN, IT
 CC SPLITS ALPHA-LINKED GALACTOSE MOIETIES. IT IS PARTICULARLY
 CC SUITABLE FOR THE HYDROLYSIS OF GUAR GUM TO A GUM WITH IMPROVED
 CC GELLING PROPERTIES. PREFERENTIALLY CLEAVES ALPHA-1,6 GLYCOSIDE
 CC LINKAGES.
 CC -!- CATALYTIC ACTIVITY: Melibiose + H(2)O -> galactose + glucose.
 CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X14619; CAA32772.1;
 DR PIR: S07472; S07472.

DR InterPro: IPR002241; Glyco_hydro_27.
 DR InterPro: IPR000111; Glyco_hydro_GHD.
 DR Pfam: PF02065; Melibiase; 1.
 DR PRINTS: PR00740; GLHYDRASE27.
 DR PRODOM: PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE: PS00512; ALPHA-GALACTOSIDASE; 1.
 KW Hydrolase; Glycosidase; Signal; Glycoprotein.
 FT SIGNAL 1 24
 FT PROPEP 25 47
 FT CHAIN 48 411 ALPHA-GALACTOSIDASE.
 FT CARBOHYD 32 32 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 145 145 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC...) (POTENTIAL).
 FT ACT_SITE 319 319 POTENTIAL.
 SQ SEQUENCE 411 AA; 45135 MW; 5B1715858D1AB1E CRC64;
 Query Match 29.3%; Score 681.5; DB 1; Length 411;
 Best Local Similarity 42.2%; Pred. No. 5.8e-52;
 Matches 141; Conservative 52; Mismatches 102; Indels 39; Gaps 9;
 QY 33 DNGLARTPTMGLWHERFMCNLDCEPDCISEKLFEMAEMLVMSGKWDAGYVLCID 92
 Db 49 ENGLGQTPPMGWSNHFHFGCD-----INENVRETADAMVSTGLAALGYQYINLD 98
 QY 93 DCWMAQPDSEGRLOADPQREPHGIRQLANYVHSGKLGKIYADVGNKTKCA-GFPGSFGY 151
 Db 99 DCWAEINROSGNVPNAAAFPSGIRKALADYVHSGKLGKIYSDAGTQCTSKTMPGSLGH 158
 QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDLSLENLA----DGYKHMSLALNRTGRSIVYS-CWPL 206
 Db 159 EEOADKTFASGWGVDLYKYDNC-----ENLGISVKERYPPMGKALLSSGRPIFFSMCE-- 210
 QY 207 YMPFKPNYTEIQVCNHNWRNADIDDSWKSISILD-----WTSNQERIVDVAGPGW 262
 Db 211 --GWEDPQTW-AKSGNSWRTTGDIEDWNWSWTSTADSNDRKASY-----AGPGCW 259
 QY 263 NDPMLVIGNFGLSNQOVTOMALWMAAPLWMSNDLRHISPOAKKALODKDVIAINQD 322
 Db 260 NDPMLVEVNGGMMTTEYRSHFSIWALAKAPLLVGCIDIRAMDDTTHELISNAEVIADVQD 319
 QY 323 PLGKQGYQLRQGNFVEWPERPLSGLAWAVAMINR 356
 Db 320 KLGVGKVKSTNDLEVMVAGPLSDNKVAVILWNR 353

RESULT 6

AGAL_ASPNG STANDARD; PRT; 545 AA.
 AC P28351;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase).
 GN AGLA.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-59.
 RC STRAIN=ATCC 9089 / N402;
 RX MEDLINE=92318893; PubMed=1320186;
 RA den Herder I.F., Rosell A.M.M., van zuijen C.M., Punt P.J.,
 RA van den Hondel C.A.M.J.J.;
 RT "Cloning and expression of a member of the Aspergillus niger gene
 RT family encoding alpha-galactosidase.";
 CC Mol. Gen. Genet. 233:404-410(1992).
 CC -!- FUNCTION: REPRESENT A MINOR EXTRACELLULAR ALPHA-GALACTOSIDASE
 CC ACTIVITY IN A NIGER.
 CC -!- CATALYTIC ACTIVITY: Melibiose + H(2)O -> galactose + glucose.
 CC -!- PTM: A C-TERMINAL SER/THR-RICH REGION MAY PROVIDE POSSIBLE SITES
 CC FOR O-GLYCOSYLATION.
 CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.


```

DR PIR; S50310; S50310.
DR SGD; L0001064; MEL2.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR006215; Glyco_hydro_36Me.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRINTS; PR00748; MELIBIASE.
DR PROSITE; PS005272; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 18
FT CHAIN 19 471
FT CARBOHYD 105 105
FT CARBOHYD 175 175
FT CARBOHYD 270 270
FT CARBOHYD 370 370
FT CARBOHYD 403 403
FT CARBOHYD 413 413
FT CARBOHYD 422 422
FT CARBOHYD 435 435
FT CARBOHYD 454 454
FT CARBOHYD 454 454
SQ SEQUENCE 471 AA; 52230 MW; 7FA36068E01F33B CRC64;

Query Match 22.0%; Score 510; DB 1; Length 471;
Best Local Similarity 32.6%; Pred. No. 6.6e-37;
Matches 122; Conservative 59; Mismatches 129; Indels 64; Gaps 10;

QY 34 NGLARTPTMGWLHWERFNCNLDQCPEPDCISEKLFEMAEMLVSEGNKDAEYELCIDDD 93
DB 24 NGLGLTPQMGWNNWNTFACD-----VSEQLLLDTADRLSLGLKDMGYKVIILDD 73

QY 94 CWMAFQRDSEGLQADPQPHGIRQLANYVHSKGLKGIYADVGNKTCAGPFGSGFYVD 153
DB 74 CW-SSGRSDGFLVADEQKFPNGMGHVADHLNNSFLFGMYSSAGEYTCAGYPSGLGRE 132

QY 154 IDAQTFADGWDLKFDGCGY-----CDSLENLADGKYKMSLALNRTGRSIVYS-CWMP--- 205
DB 133 EDAQFANNRVLYKYDNCYNKQGQGTPEISVHYRKAMSDALNKTGRPIFYSLCNWGQDL 192

QY 206 LYMPFQKPNYTEIRQYCNHNRNFADIDDSWK-----SIKSTL 243
DB 193 TFYWG-----SGIANSWRMSGDITAEFTRPDSRCPDGDDEYCKYAGFHCISIMIL 243

QY 244 DWTSEFQERIVDVAGPGGWNPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDLRHI 303
DB 244 NKAAPMGQN---AGVGWGNLDLNLVVRVGNLTDEEKAHPMAVKSPLIIGADVNTL 299

QY 304 SPQAKALLQDKDVIAINQDPLQKQG-----YOLRGDNF-----EYWERPLSGLAWAVAM 353
DB 300 KPSSYSIYSQASVIAINQDPKIPATRVVRYVYVSDTDEYGGGEIQMWSGPLDNGDQVVAL 359

QY 354 INRQEIGGPRSYTI 367
DB 360 LNGGVPRPMNTTL 373

RESULT 10
MEL1_YEAST
AC P04824;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Alpha-galactosidase 1 precursor (EC 3.2.1.22) (Melibiase 1) (Alpha-D-
DE galactoside galactohydrolase 1).
GN MEL1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_taxid:4932;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-86041913; PubMed-2997745;
RA Liljestrom P.L.;
RT "The nucleotide sequence of the yeast MEL1 gene.";
RL Nucleic Acids Res. 13:7257-7268(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Carlbergensis;
RX MEDLINE-86083175; PubMed-3000884;
RA Summer-Smith M., Bozzato R.P., Skipper N., Davies R.W., Hopper J.E.;
RT "Analysis of the inducible MEL1 gene of Saccharomyces carlsbergensis
and its secreted product, alpha-galactosidase (melibiase).";
RL Gene 36:333-340(1985).
CC -!- CATALYTIC ACTIVITY: Melibiase + H(2)O = galactose + glucose.
CC -!- SUBUNIT: Homodimer.
CC -!- INDUCTION: INDUCED BY GALACTOSE AND MELIBIOSE AND REPRESSED
CC BY GLUCOSE.
CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03102; CAA26888.1;
CC EMBL; M10604; AAA34770.1;
CC PIR; A00897; GBBYAG.
CC SGD; L0001063; MEL1.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR006215; Glyco_hydro_36Me.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRINTS; PR00748; MELIBIASE.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Multigene family; Signal.
FT SIGNAL 1 18
FT CHAIN 19 471
FT CARBOHYD 105 105
FT CARBOHYD 175 175
FT CARBOHYD 270 270
FT CARBOHYD 370 370
FT CARBOHYD 403 403
FT CARBOHYD 413 413
FT CARBOHYD 422 422
FT CARBOHYD 435 435
FT CARBOHYD 454 454
FT CARBOHYD 454 454
SQ SEQUENCE 471 AA; 52102 MW; 7A2E265A6BA09DBD CRC64;

Query Match 21.6%; Score 502; DB 1; Length 471;
Best Local Similarity 31.8%; Pred. No. 3.3e-36;
Matches 119; Conservative 63; Mismatches 128; Indels 64; Gaps 10;

QY 34 NGLARTPTMGWLHWERFNCNLDQCPEPDCISEKLFEMAEMLVSEGNKDAEYELCIDDD 93
DB 24 NGLGLTPQMGWNNWNTFACD-----VSEQLLLDTADRLSLGLKDMGYKVIILDD 73

QY 94 CWMAFQRDSEGLQADPQPHGIRQLANYVHSKGLKGIYADVGNKTCAGPFGSGFYVD 153
DB 74 CW-SSGRSDGFLVADEQKFPNGMGHVADHLNNSFLFGMYSSAGEYTCAGYPSGLGRE 132

QY 154 IDAQTFADGWDLKFDGCGY-----CDSLENLADGKYKMSLALNRTGRSIVYS-CWMP--- 205
DB 133 EDAQFANNRVLYKYDNCYNKQGQGTPEISVHYRKAMSDALNKTGRPIFYSLCNWGQDL 192

QY 206 LYMPFQKPNYTEIRQYCNHNRNFADIDDSWK-----SIKSTL 243
DB 193 TFYWG-----SGIANSWRMSGDITAEFTRPDSRCPDGDDEYCKYAGFHCISIMIL 243

QY 244 DWTSEFQERIVDVAGPGGWNPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDLRHI 303

```

```

Db 244 NKAAPMGON-----AGVGGNDLNDLEVGVCNLTDDDEKAHFSMAWVKSPLLIGANVNL 299
QY 304 SPOAKALLQDKVIAINQDPLGKQG-----YOLROGDNF-----EWERPLSLAWAVAM 353
Db 300 KASSVYISQASVIAINQDNGSIPATRVWRYVYVSDTDEYGGQGEIQMWSGPLDNGQGVVAL 359
QY 354 INRQETGGPRSYTI 367
Db 360 LNGGSVSRPMNTTL 373

RESULT 11
ID IMD_ARTGO STANDARD; PRT: 641 AA.
AC Q44052;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Isomalto-dextranase precursor (EC 3.2.1.94) (Glucan 1,6-alpha-
DE isomaltosidase) (Exo-isomaltohydrolase).
GN IMD.
OS Arthrobacter globiformis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=1665;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=T6;
RX MEDLINE=95095945; PubMed=8002600;
RA Iwai A., Ito H., Mizuno T., Mori H., Matsui H., Honma M., Okada G.,
RA Chiba S.;
RT "Molecular cloning and expression of an isomalto-dextranase gene from
RT Arthrobacter globiformis T6."
RL J. Bacteriol. 176:7730-7734(1994).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in polysaccharides so as to remove successive isomaltose units
CC from the non-reducing ends of the chains.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D30761; BAA06424.1;
CC PIR; A55549; A55549.
CC InterPro; IPR000111; Glyco_hydro_GHD.
CC InterPro; IPR006311; Tat.
CC ProDom; PD002572; Glyco_hydro_Ghd; 1.
CC TIGRFAMs; TIGR01409; TAT_signal_seq; 1.
CC PROSITE; PS00512; ALPHA_GALACTOSIDASE; FALSE_NEG.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 39
FT CHAIN 40 641
SQ SEQUENCE 641 AA; 69764 MW; B7F8F278F4D88350 CRC64;

Query Match 5.6%; Score 129; DB 1; Length 641;
Best Local Similarity 21.1%; Pred. No. 0.0019;
Matches 80; Conservative 4; Mismatches 130; Indels 126; Gaps 19;

QY 81 WK-----DAGYEYLCIDCWV--APORDSEGRLOADPQREPHGIRQLANYVHKS 127
Db 85 WKANIDWAGNFAFGYDIACDVG-WIEGSRRTGNGYITSYNDSWQHDWAYWANYLAAR 143
QY 128 GLKLGIV-----ADVGNKTCAG-----FPGSGFYDID----- 155
Db 144 KMKLGIVYNPLVWHRAAVEDASKTVLGRPDVKIADLVVPGDFEARDIGGNOLYLDVTKS 203

```

```

QY 156 -----AQTADWGVDLKFD--GCYDSLE-----NLADGYKHMSLALNFTGRSI 198
Db 204 CAKEYVOGYRYFKDGLGVYLRIDFLSWYEDGRDANIGOVNAPHGRANYELALSINEAA 263
QY 199 VYSCWPLYM-WPFO-----KPNYTEIRQYCNHWRNFADIDDSWKS 238
Db 264 GEDMEVSLVMPHMFQDGSAAELANGDLVRINADAKGGWDRLSGMRQNMW-----DAMP- 316
QY 239 IKSILDWTSFNOERIVDVAGPGWN-----DPDMLVIGFGLSWNQVQTMALWA 288
Db 317 -----NWAN-----PFCGFTGWSHRNGRCQLILDGDFMRASFTA-SDEERKTMNLMW 363
QY 289 IMAAPLEMSNDLRHISPOAKALLQDKVIAINQDPL-GKQGYQLRQGDNF-----E 338
Db 364 AAGSPLAIAADTYQOIGNNA-WVYTNKEVLQNLADGLVGKPLY--RSATPFKSCPGSRDTE 420
QY 339 VWERPLSGLAWAVAMINRQE 358
Db 421 RWAGQLPDGSGWGVLENRSD 440

RESULT 12
Y113_AQUAE STANDARD; PRT: 348 AA.
AC O66516;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_113 precursor.
GN AQ_113.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RA "The complete genome of the hyperthermophilic bacterium Aquifex
RA aeolicus."
RT Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000674; AAC06480.1;
CC PIR; A70311; A70311.
CC Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 17
FT CHAIN 18 348
FT SEQUENCE 348 AA; 38760 MW; BDDC09E0013AF152 CRC64;

Query Match 5.4%; Score 125; DB 1; Length 348;
Best Local Similarity 21.3%; Pred. No. 0.002;
Matches 77; Conservative 59; Mismatches 117; Indels 108; Gaps 21;

QY 50 EMCNLDCQEPDSCISEKLFEMAEALMVSEGNKADAGEYLCIDDCWMAFORSEGRLOAD 109
Db 10 FISSSLAE-----KSLYLVKSQPLEYGGISAGFYVTNED-----GKSSDKL--- 55
QY 110 PORPHGIRQLANYVHSGKLGKIGIYADVGNKTCAGF-----PGSGFYD 153
Db 56 -----QITNAI-----LGLKAEVGEKIRFGFDLAVGSLWSTVWDGQGFSEYD 100
QY 154 IDAQTAD-----WGVDLK-FDGCYDS-----LENADGYKHMSLALNFTGRSIYSCW 204

```

Db 101 PDEKELKEGVLHWGYAYLKPFKFEVSAGVLTNNI--GYE---LADTYSNPNTVFGAVW 155
Qy 205 PLYMWPQKPNYETIRQYCNHWRNFADIDDSKWSIKSLDWTSTNQERIVDVAGPGGWN 264
Db 156 --YAQFIYFG---VRITL-----DFSEIADLPVSL-----YAEYNQEVLDNYAVG----- 197
Qy 265 PDMVLVGNFG-----LSW-----NOQVOTOMAL-WAIIAALPLFMSNDLRHISPOAKALIQ 313
Db 198 ----ILGELGNLSVALSYDYRASKNLVDLVGLVTIENVLDGLNFDYQWLDDSAKSQD 253
Qy 314 KDVTAINQDPLGKOGYQLOD-----NFEWEPRLSLAWAVAMINRQEGGPRSYTIA 368
Db 254 DSAVGVALYVIPK--FETKYGEFSLRLEVFDEGSSGIYFG-----GADKGYTLT 302
Qy 369 V 369
Db 303 V 303

RESULT 13
AM3B_ORYSA
ID AM3B_ORYSA STANDARD; PRT; 438 AA.
AC P27937;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alpha-amylase isozyme 3B precursor (EC 3.2.1.1) (1,4-alpha-D-glucan
glucanohydrolase).
GN AMYL 6 OR AMY3B.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Japonica M202; TISSUE=Etisolated leaf;
RX MEDLINE=91329692; PubMed=1714318;
RA Sutliff T.D., Huang N., Litts J.C., Rodriguez R.L.;
RT "Characterization of an alpha-amylase multigene cluster in rice.";
RL Plant Mol. Biol. 16:579-591(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Sutliff T.D., Huang N., Rodriguez R.L.;
RL Submitted (MAY-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IMPORTANT FOR BREAKDOWN OF ENDOSPERM STARCH DURING
GERMINATION.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
CC -!- SUBUNIT: Monomer.
CC -!- TISSUE SPECIFICITY: GERMINATING SEEDS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT A HIGH LEVEL DURING GERMINATION
IN THE ALEURONES CELLS UNDER THE CONTROL OF THE PLANT HORMONE
GIBBERELIC ACID AND IN THE DEVELOPING GRAINS AT A LOW LEVEL.
CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X56337; CAA39777.1; -;
CC DR EMBL; M24941; AAA33897.1; -;
CC PIR; S14957; S14957.
CC DR HSSP; P04063; IAVA.
CC Gramene; P27937; -;
CC DR InterPro; IPR006589; Alp_aml_cat_sub.
CC DR InterPro; IPR006047; Alp_aml_cat.

DR Pfam; PF00128; alpha-amylase; 1.
DR SMART; SM00642; Amy; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Calcium; Signal;
KW Multigene family.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 438 ALPHA-AMYLASE ISOZYME 3B.
FT ACT_SITE 205 205 BY SIMILARITY.
FT ACT_SITE 313 313 BY SIMILARITY.
FT METAL 117 117 CALCIUM (BY SIMILARITY).
FT METAL 176 176 CALCIUM (BY SIMILARITY).
SQ SEQUENCE 438 AA; 48591 MW; B9DE0DB5ABC63F9C CRC64;
Query Match 4.9%; Score 114.5; DB 1; Length 438;
Best Local Similarity 25.8%; Pred. No. 0.022;
Matches 70; Conservative 30; Mismatches 96; Indels 75; Gaps 20;
Qy 63 CISEKLFMEAEMLMVS-----EGWKDAG--YEYL--CIDD-----CWM-----APQ 99
Db 18 CLSSHL--AAQVLFQGFENWESWKQGGYNFLHGHVDDIAATGVTHVWLPPLPPSHVAPQ 75
Qy 100 RDEGRLL-QADPQRFPHG--IROLANYVHSKGLKGIYAD-VGNKTCAGFGPGSGFYDI- 154
Db 76 GYMGRLYDLDAKSYGTGAEDRSLLAAPHSGIK--CVADIVINHRADYKDSRGYICIF 133
Qy 155 ---DAQTFADMGVLLKFDGCYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLYMMPF 211
Db 134 EGGTPDSRLDWDGPDWI-----CSDDTQYSNGRHRD-----TGADFGAA-----PD 174
Qy 212 QKPNYTEIRQYCNHWRNFADID---DSWKSITSILD-----WTSFNQERIVDVAGPGG--- 261
Db 175 IDHLNTRVQTELSDLNWLKSDVGFQDGNR-----LDFAKGYSAAVAKTYVDNTDPSFVVA 229
Qy 262 --WNDPDMVLVGNFGLSNQVQTQMAL--WA 288
Db 230 EIWS--NMRYDNGEPEPSWNQDGRQELVNW 258

RESULT 14
AGAL_PEPDE
ID AGAL_PEPDE STANDARD; PRT; 733 AA.
AC P43467;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Alpha-galactosidase 1 (EC 3.2.1.22) (Melibiase).
GN AGAR.
OS Pedicoccus pentosaceus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pediococcus.
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PPeL.0;
RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ALPHA-GALACTOSIDASE ASSOCIATED WITH THE RAFFINOSE
OPERON.
CC -!- CATALYTIC ACTIVITY: Melibiose + H(2)O -> galactose + glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 36 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z32771; CAA83665.1; -;
CC DR EMBL; L32093; AA25564.1; -;
CC PIR; S44254; S44254.
CC DR InterPro; IPR002252; Glyco_hydro_36.
CC DR InterPro; IPR000111; Glyco_hydro_GHD.
CC Pfam; PF02065; Melibiase; 1.

DR PRINTS; PR00743; GLHYDRIASE36.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 KW Hydrolase; Glycosylase; Multigene family.
 SQ SEQUENCE. 733 AA; 83313 MW; 7732ACFDB655EE88 CRC64;

Query Match 4.9%; Score 114; DB 1; Length 733;
 Best Local Similarity 21.1%; Pred. No. 0.046;
 Matches 90; Conservative 43; Mismatches 148; Indels 146; Gaps 21;

QY 66 EKLFEMAEEL---MVSEGWKAGYEYLCIDCWAPORD---SEGRLOADPQRPFGH 118
 DB 339 EATFFDFNEAKLPITVDEA-KOLGTEMFVLDGWFGRHDDNSSLGWDQVDRKFPQGLN 397
 QY 119 QLANYVSHKGLKGIY-----ADVGNKTC 142
 DB 398 HFVKYVHSGKLFGLWLEPEMISYDSKLYQQHPDYLMOVGRSPSRNQVLDLGRQAV 457
 QY 143 AGFGSGFYDIDQTFADMGVDLLKFD-----GCYCD 175
 DB 458 RN-----NIFQDLQLLKSKOIDYIKWDMNRHLSIYSVALPPERQGEVYHRYVGLY-E 511
 QY 176 SLENLADCYKMSALNLTGRSIVYSCWPLYMPPFQKPNYTEIRQCNHWRNFADIDDS 235
 DB 512 LLERLTAYPHI-LFEGCGGGGRFDAGMAYM-----PQI----- 546
 QY 236 WKSISILDTWSFNOERTVDVAGPGWN--DPDMLVIGNGLSNQOV---TOMALNIM 290
 DB 547 -----WASDNTDAVARLTIOYGLSLAPISLATAHVSVPNQQTGRTSMTSAV 597
 QY 291 AAP-LFMSNDRHISPOAKALLQKDVIAINQ-DPLGKQG--YOLRQGDNFVWERPLS 345
 DB 598 AASGLVGLYELDTLSSADKQIVQ-KOVVOYKQIRLTIQFGEFYRLKS-----PIT 647
 QY 346 G-LAWAVAMINRQE---IGGPRSYTIAVASLCKGVACNPACFITOLLPVK----- 391
 DB 648 SNQRAWFEVSPQDEAIVWFNLTSYAQPSLTKTGLVLPK-LYONIAATKATFGGDEL 706
 QY 392 RKLGFYE 398
 DB 707 MOLGYFD 713

RESULT 15

RAFA_ECOLI STANDARD; PRT; 708 AA.
 AC PI6551;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Alpha-galactosidase (EC 3.2.1.22) (Mellibiose).
 GN RAFA.
 OS Escherichia coli.
 OG Plasmid PRSD2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90078124; PubMed=2556373;
 RA Aslanidis C., Schmid K., Schmitt R.;
 RT "Nucleotide sequences and operon structure of plasmid-borne genes
 mediating uptake and utilization of raffinose in Escherichia coli.";
 RL J. Bacteriol. 171:6753-6763(1989).
 RN [2]
 RP SEQUENCE OF 1-6 FROM N.A.
 RX MEDLINE=90202743; PubMed=2180920;
 RA Aslanidis C., Schmitt R.;
 RT "Regulatory elements of the raffinose operon: nucleotide sequences of
 operator and repressor genes.";
 RL J. Bacteriol. 172:2178-2180(1990).
 RN [3]

RP SEQUENCE OF 1-5.
 RX MEDLINE=77003028; PubMed=786627;
 RA Schmid K., Schmitt R.;
 RT "Raffinose metabolism in Escherichia coli K12. Purification and
 properties of a new alpha-galactosidase specified by a transmissible
 plasmid.";
 RL Eur. J. Biochem. 67:95-104(1976).
 CC -!- CATALYTIC ACTIVITY: Melibiose + H(2)O = galactose + glucose.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- SIMILARITY: BELONGS TO FAMILY 36 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M27273; AAA24497.1; --
 CC EMBL; M29849; AAA24501.1; --
 CC PIR; A43717; A43717.
 CC InterPro; IPR002252; Glyco_hydro_36.
 CC InterPro; IPR000111; Glyco_hydro_GHD.
 CC Pfam; PF02065; Melibiase; 1.
 CC PRINTS; PR00743; GLHYDRIASE36.
 CC PRODOM; PD002572; Glyco_hydro_GHD; 1.
 CC PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 KW Hydrolase; Glycosidase; Plasmid.
 SQ SEQUENCE 708 AA; 81188 MW; 3D915DA5DB0A202E CRC64;

Query Match 4.3%; Score 100.5; DB 1; Length 708;
 Best Local Similarity 23.9%; Pred. No. 0.67;
 Matches 37; Conservative 16; Mismatches 41; Indels 61; Gaps 8;

QY 24 WDIPGARAL-DNGLRATPTMGWLH-----WLYACHSADGLNGMSQQYHRLRDEIRFEQKLRPVHL 296
 DB 241 W-MPGEKALKRKNETLVTP---WLYACHSADGLNGMSQQYHRLRDEIRFEQKLRPVHL 296
 QY 47 --WERFMCNLDCEEPDSCISEKLFEMAEYLVSEGWKAGYEYLCIDCWAPQRDSEG 104
 DB 297 NTWEGYFN-----HNFD-----YIMQAEAAAA-----LGVERFTIDGWFGRNDRA 341
 QY 105 RL---QADPQRPFGHTRQLANYVSHKGLKGIYAD 136
 DB 342 ALGDWYTDQKYPNGLMPVINVHVKSLGMEFGIWE 376

Search completed: July 29, 2003, 16:15:00
 Job time : 29.4966 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:09:29 ; Search time 38.7987 Seconds
(without alignments)
461.292 Million cell updates/sec

Title: US-09-993-059-18
Perfect score: 2322
Sequence: 1 MOLRNPELHLCALALFLA.....RSHINPTGTVLQLSEKDEL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2293	98.8	429	1	US-07-602-824A-2
2	2293	98.8	429	1	US-07-602-608-3
3	2293	98.8	429	1	US-07-983-451-2
4	2293	98.8	429	1	US-08-261-578-3
5	2293	98.8	429	1	US-08-261-577-7
6	2293	98.8	429	1	US-08-261-577-9
7	2289	98.6	429	3	US-09-070-356-4
8	2275	98.0	428	6	5179023-4
9	2131	91.8	387	3	US-09-176-666-11
10	2131	91.8	387	3	US-09-176-666-10
11	2131	91.8	388	3	US-09-176-666-9
12	2131	91.8	389	3	US-09-176-666-8
13	2131	91.8	390	3	US-09-176-666-7
14	2131	91.8	391	3	US-09-176-666-6
15	2131	91.8	392	3	US-09-176-666-5
16	2131	91.8	393	3	US-09-176-666-4
17	2131	91.8	394	3	US-09-176-666-3
18	2131	91.8	395	3	US-09-176-666-2
19	2131	91.8	398	3	US-08-928-881-26
20	2131	91.8	398	3	US-09-176-666-1
21	2131	91.8	398	4	US-09-543-921-26
22	2131	91.8	398	4	US-09-266-014-4
23	2131	91.8	398	4	US-09-491-759-26
24	2110	90.9	381	3	US-09-176-666-12
25	1061	45.7	406	3	US-09-070-356-2
26	1054	45.4	405	1	US-08-406-070-2
27	952.5	41.0	411	1	US-07-602-608-2

28	952.5	41.0	411	1	US-08-261-578-2	Sequence 2, Appli
29	952.5	41.0	411	3	US-09-070-356-3	Sequence 3, Appli
30	945.5	40.7	411	1	US-07-602-824A-3	Sequence 3, Appli
31	945.5	40.7	411	1	US-07-983-451-3	Sequence 3, Appli
32	945.5	40.7	411	1	US-08-261-577-8	Sequence 8, Appli
33	710.5	30.6	420	4	US-08-113-890-2	Sequence 2, Appli
34	687.5	29.6	363	3	US-08-488-961-4	Sequence 4, Appli
35	687.5	29.6	363	3	US-08-973-297-4	Sequence 4, Appli
36	687.5	29.6	363	5	PCT-US96-06511-4	Sequence 4, Appli
37	680.5	29.3	363	1	US-08-488-961-7	Sequence 7, Appli
38	680.5	29.3	363	3	US-08-973-297-7	Sequence 7, Appli
39	680.5	29.3	363	5	PCT-US96-06511-7	Sequence 7, Appli
40	679.5	29.3	411	3	US-09-070-356-6	Sequence 6, Appli
41	600	25.8	404	1	US-07-602-824A-4	Sequence 4, Appli
42	600	25.8	404	1	US-07-602-608-4	Sequence 4, Appli
43	600	25.8	404	1	US-07-983-451-4	Sequence 4, Appli
44	600	25.8	404	1	US-08-261-578-4	Sequence 4, Appli
45	600	25.8	404	1	US-08-261-577-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1
US-07-602-824A-2
; Sequence 2, Application US/07602824A
; Patent No. 5356804
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE alpha-GALACTOSIDASE A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602.824A
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-2

Query Match 98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.9e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHLCALALFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDQKEP 60
DB 1 MOLRNPELHLCALALFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLDQKEP 60


```
Query Match          98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.9e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNCNDQCEEP 60
DB 1 MQLRNPGLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNCNDQCEEP 60
QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAQPDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAQPDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKLGIIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
DB 121 ANYVHSKGLKLGIIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
QY 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSENOBRIVDVAGPGGWNDDMLVIGNFGLSWNOQVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSENOBRIVDVAGPGGWNDDMLVIGNFGLSWNOQVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTTAVASLGLKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTTAVASLGLKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 4
US-08-261-578-3
; Sequence 3, Application US/08261578
; Patent No. 5491075
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,578
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/602,608
; FILING DATE: 24-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-261-578-3

Query Match          98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.9e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNCNDQCEEP 60
DB 1 MQLRNPGLHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNCNDQCEEP 60
QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAQPDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAQPDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKLGIIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
DB 121 ANYVHSKGLKLGIIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
QY 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSENOBRIVDVAGPGGWNDDMLVIGNFGLSWNOQVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSENOBRIVDVAGPGGWNDDMLVIGNFGLSWNOQVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTTAVASLGLKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTTAVASLGLKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 5
US-08-261-577-7
; Sequence 7, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; REFERENCE/DOCKET NUMBER: 6923-042
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-577-7

Query Match 98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.9e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
DB 1 MOLRNPELHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
QY 61 DSCISEKLFEMAEMLVSEGWKDAYEYLCIDDCWMAQPDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFEMAEMLVSEGWKDAYEYLCIDDCWMAQPDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYGYDIDAQTFADWGVLLKFDGVCYCDLSLE 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYGYDIDAQTFADWGVLLKFDGVCYCDLSLE 180
QY 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHNRNFADIDDSWSIK 240
DB 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHNRNFADIDDSWSIK 240
QY 241 SILDWTSEFQERIVDVAGPGGWNDDMLVIGNGLSWNQVOTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSEFQERIVDVAGPGGWNDDMLVIGNGLSWNQVOTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVIAINQDPLGKQYQIRQDGFVEWRLPSGLAWAVAMINROEIG 360
DB 301 RHISPOKALLQDKVIAINQDPLGKQYQIRQDGFVEWRLPSGLAWAVAMINROEIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGEYEWTSRLRSHINPTGTVLLQL 417
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGEYEWTSRLRSHINPTGTVLLQL 417

RESULT 6
US-08-261-577-9
Sequence 9, Application US/08261577
Patent No. 5580757
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-Galactosidase A
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,577
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-042
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-261-577-9

Query Match 98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.9e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
DB 1 MOLRNPELHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
QY 61 DSCISEKLFEMAEMLVSEGWKDAYEYLCIDDCWMAQPDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFEMAEMLVSEGWKDAYEYLCIDDCWMAQPDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYGYDIDAQTFADWGVLLKFDGVCYCDLSLE 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYGYDIDAQTFADWGVLLKFDGVCYCDLSLE 180
QY 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHNRNFADIDDSWSIK 240
DB 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHNRNFADIDDSWSIK 240
QY 241 SILDWTSEFQERIVDVAGPGGWNDDMLVIGNGLSWNQVOTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSEFQERIVDVAGPGGWNDDMLVIGNGLSWNQVOTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVIAINQDPLGKQYQIRQDGFVEWRLPSGLAWAVAMINROEIG 360
DB 301 RHISPOKALLQDKVIAINQDPLGKQYQIRQDGFVEWRLPSGLAWAVAMINROEIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGEYEWTSRLRSHINPTGTVLLQL 417
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGEYEWTSRLRSHINPTGTVLLQL 417

RESULT 7
US-09-070-356-4
Sequence 4, Application US/09070356
Patent No. 6228631
GENERAL INFORMATION:
APPLICANT: Alex Zhu
APPLICANT: Jack Goldstein
TITLE OF INVENTION: Recombinant a-N-
TITLE OF INVENTION: Acetylglucosaminidase
TITLE OF INVENTION: Enzyme and cDNA Encoding
TITLE OF INVENTION: Said Enzyme
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amster, Rothstein & Ebenstein
STREET: 90 Park Avenue
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,356
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,248
FILING DATE: March 26, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 63475/12
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 429
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: cDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: yes
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: library
POSITION IN GENOME: unknown
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY: human a-galactosidase
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Calhoun et al
TITLE: Fabry Disease: Isolation of a cDNA
TITLE: Clone Encoding Human a-Galactosidase A
JOURNAL: Proceedings of the National Academy
JOURNAL: of Science USA
VOLUME: 82
PAGES: 7364-7368
DATE: 1985
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

Query Match 98.6%; Score 2289; DB 3; Length 429;
Best Local Similarity 99.0%; Pred. No. 7.7e-232;
Matches 413; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALREFALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLRNPDLHLCALALREFALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFDGICYDLSLENL 180

Db 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFDGICYDLSLENL 180
Qy 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSENOERIVDVAGPGGWNPDMLVIGNFGLSNQOQVTOMALWAIMAPLFMSNDL 300
Db 241 SILDWTSENOERIVDVAGPGGWNPDMLVIGNFGLSNQOQVTOMALWAIMAPLFMSNDL 300
Qy 301 RHISPAKALLODKDVIATINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINROEIG 360
Db 301 RHISPAKALLODKDVIATINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVVRKKGFEYWTSLRSHINPTGTVLLOL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVVRKKGFEYWTSLRSHINPTGTVLLOL 417
RESULT 8
5179023-4
Patent No. 5179023
APPLICANT: CALHOUN, DAVID H.; COPPOLA, GEORGE
TITLE OF INVENTION: RECOMBINANT a-GALACTOSIDASE A THERAPY
FOR FABRY DISEASE
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/520,312
FILING DATE: 07-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 328,421
FILING DATE: 24-MAR-1989
SEQ ID NO: 4
LENGTH: 428
5179023-4

Query Match 98.0%; Score 2275; DB 6; Length 428;
Best Local Similarity 99.3%; Pred. No. 2.3e-230;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALREFALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLRNPDLHLCALALREFALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFDGICYDLSLENL 180
Db 121 ANYVHSGKLGIGYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFDGICYDLSLENL 180
Qy 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXKMSLALNRTGRSIVSCEWPLYMWPQKPNYTEIROYNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSENOERIVDVAGPGGWNPDMLVIGNFGLSNQOQVTOMALWAIMAPLFMSNDL 300
Db 241 SILDWTSENOERIVDVAGPGGWNPDMLVIGNFGLSNQOQVTOMALWAIMAPLFMSNDL 300
Qy 301 RHISPAKALLODKDVIATINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINROEIG 360
Db 301 RHISPAKALLODKDVIATINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVVRKKGFEYWTSLRSHINPTGTVLLOL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVVRKKGFEYWTSLRSHINPTGTVLLOL 417
RESULT 9
US-09-176-666-11
Sequence 11, Application US/09176666

```
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-11

Query Match
Best Local Similarity 91.8%; Score 2131; DB 3; Length 386;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 91
Db 1 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 60
QY 92 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 211
Db 121 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 180
QY 212 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 271
Db 181 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 240
QY 272 NFGLSWNOQVTOALWAIMAAPLMSNDLRHISPOAKALLODKDVIAINODPLGKQGYQL 331
Db 241 NFGLSWNOQVTOALWAIMAAPLMSNDLRHISPOAKALLODKDVIAINODPLGKQGYQL 300
QY 332 RQGNFVWERPLSGLAWAVAMINROEIGGPRSYTTAVASLKGKGVACNPACFITOLLPVK 391
Db 301 RQGNFVWERPLSGLAWAVAMINROEIGGPRSYTTAVASLKGKGVACNPACFITOLLPVK 360
QY 392 RKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 RKLGFYEWTSRLRSHINPTGTVLLQL 386

RESULT 10
US-09-176-666-10
; Sequence 10, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-10

Query Match
Best Local Similarity 91.8%; Score 2131; DB 3; Length 387;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 91
Db 1 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 60
QY 92 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 211
Db 121 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 180
QY 212 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 271
Db 181 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 240

Query Match
Best Local Similarity 91.8%; Score 2131; DB 3; Length 387;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 91
Db 1 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 60
QY 92 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 211
Db 121 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 180
QY 212 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 271
Db 181 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 240
```

```
Best Local Similarity 100.0%; Pred. No. 2.6e-215;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 91
Db 1 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 60
QY 92 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 211
Db 121 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 180
QY 212 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 271
Db 181 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 240
QY 272 NFGLSWNOQVTOALWAIMAAPLMSNDLRHISPOAKALLODKDVIAINODPLGKQGYQL 331
Db 241 NFGLSWNOQVTOALWAIMAAPLMSNDLRHISPOAKALLODKDVIAINODPLGKQGYQL 300
QY 332 RQGNFVWERPLSGLAWAVAMINROEIGGPRSYTTAVASLKGKGVACNPACFITOLLPVK 391
Db 301 RQGNFVWERPLSGLAWAVAMINROEIGGPRSYTTAVASLKGKGVACNPACFITOLLPVK 360
QY 392 RKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 RKLGFYEWTSRLRSHINPTGTVLLQL 386

RESULT 11
US-09-176-666-9
; Sequence 9, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-9

Query Match
Best Local Similarity 100.0%; Pred. No. 2.6e-215;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 91
Db 1 LDNGLARTPTMGWLHWFRCNLDCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCI 60
QY 92 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRQLADPQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 211
Db 121 YDIDAQTFADGWGVDLLKFDGCGYCDSENLDGKYMALNRTGRSIVYSCWPLYMWP 180
QY 212 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 271
Db 181 QKPNYTEIROYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGWNDPDLVIG 240
```


QY 272 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 331
DB 241 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 300
QY 332 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 391
DB 301 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 360
QY 332 RKLGFYEWTSLRSHINPTGTVLLQL 417
DB 361 RKLGFYEWTSLRSHINPTGTVLLQL 386

RESULT 12
US-09-176-666-8
; Sequence 8, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-8

Query Match 91.8%; Score 2131; DB 3; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.6e-215; Indels 0; Gaps 0;
Matches 386; Conservative 0; Mismatches 0;

QY 32 LDNGLARTPTMGWLHWFRCMNCDCQEEPDSCISEKLFEMAEALMVSEGKWDAGYEYLCI 91
DB 1 LDNGLARTPTMGWLHWFRCMNCDCQEEPDSCISEKLFEMAEALMVSEGKWDAGYEYLCI 60
QY 92 DDCWAPORDESEGRLOADPQRPFGHGIROLANYVHSGKGLGIYADVGNKTCAGFPGSFGY 151
DB 61 DDCWAPORDESEGRLOADPQRPFGHGIROLANYVHSGKGLGIYADVGNKTCAGFPGSFGY 120
QY 152 YDIDAQTFADMGVDLLKFDGCVCDLSLENLADGYKMSLALNRTGSRIVYSCWPLYMMPF 211
DB 121 YDIDAQTFADMGVDLLKFDGCVCDLSLENLADGYKMSLALNRTGSRIVYSCWPLYMMPF 180
QY 212 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNOERIVDVAGPGWNDPDMVLVG 271
DB 181 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNOERIVDVAGPGWNDPDMVLVG 240
QY 272 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 331
DB 241 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 300
QY 332 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 391
DB 301 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 360
QY 332 RKLGFYEWTSLRSHINPTGTVLLQL 417
DB 361 RKLGFYEWTSLRSHINPTGTVLLQL 386

RESULT 13
US-09-176-666-7
; Sequence 7, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro

; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-7

Query Match 91.8%; Score 2131; DB 3; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.7e-215; Indels 0; Gaps 0;
Matches 386; Conservative 0; Mismatches 0;

QY 32 LDNGLARTPTMGWLHWFRCMNCDCQEEPDSCISEKLFEMAEALMVSEGKWDAGYEYLCI 91
DB 1 LDNGLARTPTMGWLHWFRCMNCDCQEEPDSCISEKLFEMAEALMVSEGKWDAGYEYLCI 60
QY 92 DDCWAPORDESEGRLOADPQRPFGHGIROLANYVHSGKGLGIYADVGNKTCAGFPGSFGY 151
DB 61 DDCWAPORDESEGRLOADPQRPFGHGIROLANYVHSGKGLGIYADVGNKTCAGFPGSFGY 120
QY 152 YDIDAQTFADMGVDLLKFDGCVCDLSLENLADGYKMSLALNRTGSRIVYSCWPLYMMPF 211
DB 121 YDIDAQTFADMGVDLLKFDGCVCDLSLENLADGYKMSLALNRTGSRIVYSCWPLYMMPF 180
QY 212 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNOERIVDVAGPGWNDPDMVLVG 271
DB 181 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNOERIVDVAGPGWNDPDMVLVG 240
QY 272 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 331
DB 241 NFGLSWNOQVOTOMALWATMAAPLFMSNDLRHISPOAKALLQDKVIAINQDPLGKQGYOL 300
QY 332 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 391
DB 301 ROGDNFEWVERPLSLAWAVAMINRQETGGPRSYTIAVASLGKGVACNCPACFITOLLPVK 360
QY 332 RKLGFYEWTSLRSHINPTGTVLLQL 417
DB 361 RKLGFYEWTSLRSHINPTGTVLLQL 386

RESULT 14
US-09-176-666-6
; Sequence 6, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-6

Query Match 91.8%; Score 2131; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.7e-215; Indels 0; Gaps 0;
Matches 386; Conservative 0; Mismatches 0;

```
QY 32 LDNGLARTPTMGWLHWRFCMCNLDCEPDSCISEKLFEMAEMLVSEGWKADAGEYELCI 91
Db 1 LDNGLARTPTMGWLHWRFCMCNLDCEPDSCISEKLFEMAEMLVSEGWKADAGEYELCI 60
QY 92 DDCWMAPORDSEGRLOADPQRPFGHIGRLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRLOADPQRPFGHIGRLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWDLKFDGCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLYMWPF 211
Db 121 YDIDAQTFADGWDLKFDGCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLYMWPF 180
QY 212 QKPNYTEIROYNCHNRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGGWNDDPMLVIG 271
Db 181 QKPNYTEIROYNCHNRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGGWNDDPMLVIG 240
QY 272 NFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINQDPLGKQGYOL 331
Db 241 NFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINQDPLGKQGYOL 300
QY 332 ROGNFEWVERPLSGLAWAVAMINROETIGGPRSYTIAVASLGKGVACNPACFITQLLPVK 391
Db 301 ROGNFEWVERPLSGLAWAVAMINROETIGGPRSYTIAVASLGKGVACNPACFITQLLPVK 360
QY 392 RKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 RKLGFYEWTSLRSHINPTGTVLLQL 386
```

RESULT 15

```
US-09-176-666-5
; Sequence 5, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176.666
; EARLIER FILING DATE: 1998-10-21
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-5
```

```
Query Match 91.8%; Score 2131; DB 3; Length 392;
Best Local Similarity 100.0%; Pred. No. 2.7e-215;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 LDNGLARTPTMGWLHWRFCMCNLDCEPDSCISEKLFEMAEMLVSEGWKADAGEYELCI 91
Db 1 LDNGLARTPTMGWLHWRFCMCNLDCEPDSCISEKLFEMAEMLVSEGWKADAGEYELCI 60
QY 92 DDCWMAPORDSEGRLOADPQRPFGHIGRLANYVHSGKLGKIYADVGNKTCAGPFGSGY 151
Db 61 DDCWMAPORDSEGRLOADPQRPFGHIGRLANYVHSGKLGKIYADVGNKTCAGPFGSGY 120
QY 152 YDIDAQTFADGWDLKFDGCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLYMWPF 211
Db 121 YDIDAQTFADGWDLKFDGCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLYMWPF 180
QY 212 QKPNYTEIROYNCHNRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGGWNDDPMLVIG 271
Db 181 QKPNYTEIROYNCHNRNFADIDDSWKSISILDTWTSFNOERIVDVAGPGGWNDDPMLVIG 240
QY 272 NFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINQDPLGKQGYOL 331
Db 241 NFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINQDPLGKQGYOL 300
```

```
QY 332 ROGNFEWVERPLSGLAWAVAMINROETIGGPRSYTIAVASLGKGVACNPACFITQLLPVK 391
Db 301 ROGNFEWVERPLSGLAWAVAMINROETIGGPRSYTIAVASLGKGVACNPACFITQLLPVK 360
QY 392 RKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 RKLGFYEWTSLRSHINPTGTVLLQL 386
```

Search completed: July 29, 2003, 16:17:57
Job time : 39.7987 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 15:40:04 ; Search time 114.503 Seconds
(without alignments)
586.370 Million cell updates/sec

Title: US-09-993-059-18

Perfect score: 2322

Sequence: 1 MQRNPELHLCALALFLA.....RSHNPCTVLLQLSEKDEL 423

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2322	100.0	423	23	AAE28213		Human rGAL-12R pro
2	2310	99.5	427	23	AAE28211		Human rGAL-8R prot
3	2308	99.4	431	23	AAE28209		Human rGAL-4R prot
4	2306	99.3	435	23	AAE28207		Human WT rGAL-AR p
5	2293	98.8	417	23	AAE28212		Human rGAL-12 prot
6	2293	98.8	421	23	AAE28210		Human rGAL-8 prote
7	2293	98.8	429	15	AAE28234		Human alpha-galact
8	2293	98.8	429	15	AAE28234		Alpha-galactosidas
9	2293	98.8	429	16	AAE28207		Alpha-galactosidas

10	2293	98.8	429	23	AAE28206		Human WT rGAL-A pr
11	2293	98.8	478	20	AAE28206		Human breast tumou
12	2290	98.6	429	11	AAE28206		Precursor form of
13	2288	98.5	424	23	AAE28208		Human rGAL-4 prote
14	2282	98.3	429	18	AAE28201		Alpha-galactosidas
15	2275	98.0	428	14	AAE28201		Human alpha-galact
16	2234.5	96.2	409	23	AAE28215		Human rGAL-25R pro
17	2213	95.3	401	23	AAE28214		Human rGAL-25 prot
18	2165	93.2	430	24	ABG74455		Human alpha-galact
19	2131	91.8	386	22	ABG74455		Human alpha-galact
20	2131	91.8	387	22	ABG74455		Mutant delta 12 al
21	2131	91.8	388	22	ABG74455		Mutant delta 11 al
22	2131	91.8	389	22	ABG74455		Mutant delta 10 al
23	2131	91.8	390	22	ABG74455		Mutant delta 9 alp
24	2131	91.8	391	22	ABG74455		Mutant delta 8 alp
25	2131	91.8	392	22	ABG74455		Mutant delta 7 alp
26	2131	91.8	393	22	ABG74455		Mutant delta 6 alp
27	2131	91.8	394	22	ABG74455		Mutant delta 5 alp
28	2131	91.8	396	22	ABG74455		Mutant delta 4 alp
29	2131	91.8	398	22	ABG74455		Mutant delta 2 alp
30	2131	91.8	398	21	ABG74455		Amino acid sequenc
31	2131	91.8	398	21	ABG74455		Human alpha-Gal A
32	2131	91.8	398	22	ABG74455		Amino acid sequenc
33	2131	91.8	398	22	ABG74455		Alpha-galactosidas
34	2110	90.9	381	22	ABG74455		Human alpha-gal A
35	1061	45.7	406	18	AAE28204		Mutant delta 17 al
36	1060	45.7	406	15	AAE28204		Chicken liver alph
37	1054	45.4	405	18	AAE28204		Chicken liver alph
38	984	42.4	386	22	AAE28204		Chicken alpha-N-ac
39	967.5	41.7	384	22	AAE28204		Drosophila melanog
40	952.5	41.0	411	13	AAE28204		Drosophila melanog
41	952.5	41.0	411	13	AAE28204		Alpha-GalNAc from
42	952.5	41.0	411	15	AAE28204		Alpha-N-acetylglala
43	945.5	40.7	411	23	AAE28204		Human N-acetylglala
44	930	40.1	359	22	AAE28204		Alpha-galactosidas
45	714	30.7	378	16	AAE28204		Novel human secret
							Alpha-galactosidas

ALIGNMENTS

RESULT 1
AAE28213
ID AAE28213 standard; Protein; 423 AA.
AC AAE28213;
DT 27-DEC-2002 (first entry)
XX Human rGAL-12R protein.
DE Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-12R.
XX Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX US2002088024-A1.
PN 04-JUL-2002.
PD 13-NOV-2001; 2001US-0993059.
XX 26-JUL-2000; 2000US-0626127.
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
PI

XX WPI: 2002-681656/73.
DR N-PSDB; AAD45226.
XX Novel human alpha-galactosidase polypeptide useful for treating
PT lysosomal storage diseases
XX Claim 7; Page 51-52; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing
CC to a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is: human rGAL-12R protein.
XX Sequence 423 AA;
SQ

Query Match 100.0%; Score 2322; DB 23; Length 423;
Best Local Similarity 100.0%; Pred. No. 7e-223;
Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCNLDQCEEP 60
DB 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCNLDQCEEP 60
OY 61 DSCISEKLFEMAEALMVSEGWKDAYEYLICIDCWMAPOQDSEGRLOADPQRPFGHIGIOL 120
DB 61 DSCISEKLFEMAEALMVSEGWKDAYEYLICIDCWMAPOQDSEGRLOADPQRPFGHIGIOL 120
OY 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCGYCDSELENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCGYCDSELENL 180
OY 181 ADGKHMSLALNRTGRSIVISCEWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKHMSLALNRTGRSIVISCEWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
OY 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
OY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWEPRLSLGLAWAVAMINRQEIIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWEPRLSLGLAWAVAMINRQEIIG 360
OY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 420
OY 421 DEL 423
DB 421 DEL 423

RESULT 2
AAE28211
ID AAE28211 standard; Protein: 427 AA.
XX AAE28211;
XX AAE28211;
XX 27-DEC-2002 (first entry)
XX Human rGAL-8R protein.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8R.
XX

OS Homo sapiens.
XX Key Location/Qualifiers
FH Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX US2002088024-A1.
XX 04-JUL-2002.
XX 13-NOV-2001; 2001US-0993059.
XX 26-JUL-2000; 2000US-0626127.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI: 2002-681656/73.
XX N-PSDB; AAD45224.
XX Novel human alpha-galactosidase polypeptide useful for treating
PT lysosomal storage diseases
XX Claim 7; Page 46-47; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing
CC to a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-8R protein.
XX Sequence 427 AA;
SQ
Query Match 99.5%; Score 2310; DB 23; Length 427;
Best Local Similarity 99.1%; Pred. No. 1.1e-221;
Matches 423; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
OY 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCNLDQCEEP 60
DB 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCNLDQCEEP 60
OY 61 DSCISEKLFEMAEALMVSEGWKDAYEYLICIDCWMAPOQDSEGRLOADPQRPFGHIGIOL 120
DB 61 DSCISEKLFEMAEALMVSEGWKDAYEYLICIDCWMAPOQDSEGRLOADPQRPFGHIGIOL 120
OY 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCGYCDSELENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCGYCDSELENL 180
OY 181 ADGKHMSLALNRTGRSIVISCEWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKHMSLALNRTGRSIVISCEWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
OY 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQOQVQTMALWAIMAAPLFMSNDL 300
OY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWEPRLSLGLAWAVAMINRQEIIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWEPRLSLGLAWAVAMINRQEIIG 360
OY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 417
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 420

CC		produce more effective therapeutic enzyme, for producing antibodies against lysosomal enzymes for medical diagnostic use, and in any commercial process that involves substrate hydrolysis. The present sequence is human WT rGAL-AR protein.
XX		
QQ	Sequence	435 AA;
<hr/>		
	Query Match	99.3%; Score 2306; DB 23; Length 435;
	Best Local Similarity	97.2%; Pred. No. 2.9e-21;
	Matches 423; Conservative	0; Mismatches 0; Indels 12; Gaps 1;
<hr/>		
QY	1	MQLRNPETHLGGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCEEP 60
DB	1	MQLRNPETHLGGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCEEP 60
<hr/>		
QY	61	DSCISEKLFMEAEIMVSEGWKDAGYEYLICIDDCWMAQRDSERGLQADPQRFPHGIROL 120
DB	61	DSCISEKLFMEAEIMVSEGWKDAGYEYLICIDDCWMAQRDSERGLQADPQRFPHGIROL 120
<hr/>		
QY	121	ANYVHSKGLKLGIYADVGNKTCAGFGSGFYGDIDAOTFADWGVDLLKFDGCYCDSLLENL 180
DB	121	ANYVHSKGLKLGIYADVGNKTCAGFGSGFYGDIDAOTFADWGVDLLKFDGCYCDSLLENL 180
<hr/>		
QY	181	ADGYKHMSLAALNRGTGRSIVYSCEWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWSIK 240
DB	181	ADGYKHMSLAALNRGTGRSIVYSCEWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWSIK 240
<hr/>		
QY	241	SILDWTSFNOERIVDVAGPGGWNDPDMLVIGNFLGSWNQOVTOMALWAIAAPLAFNSNDL 300
DB	241	SILDWTSFNOERIVDVAGPGGWNDPDMLVIGNFLGSWNQOVTOMALWAIAAPLAFNSNDL 300
<hr/>		
QY	301	RHTSPQAKALLQDKVIATINQDPGLKGQYQLRGQDNFEVWERPLSLGLAWAVAMINROEIG 360
DB	301	RHTSPQAKALLQDKVIATINQDPGLKGQYQLRGQDNFEVWERPLSLGLAWAVAMINROEIG 360
<hr/>		
QY	361	GPRSYTIAVASLGKGVACNPACFITOLLPVKRKLGFEYWTSLRSHINPTGTVLLO --- 416
DB	361	GPRSYTIAVASLGKGVACNPACFITOLLPVKRKLGFEYWTSLRSHINPTGTVLLO --- 416
<hr/>		
QY	417	-----LSEKDEL 423
DB	421	MQMSLKDLLSEKDEL 435
<hr/>		
RESULT 5		
AAE28212	ID	AAE28212 standard; Protein; 417 AA.
XX	XX	
AC	AAE28212;	
DT	27-DEC-2002	(first entry)
XX	XX	
DE	Human	rGAL-12 protein.
KX	Human;	alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW	therapeutic;	rGAL-12.
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
FT	Cleavage-site	401..402
FT	/note=	"CTPP cleavage site"
XX	US2002088024-A1.	
XX	04-JUL-2002.	
XX	13-NOV-2001;	2001US-0993059.
XX	26-JUL-2000;	2000US-0626127.
PA	(GARG/) GARGER S J.	
PA	(TURP/) TURPEN T H.	

PA	(KUMA/) KUMAGAI M H.
XX	Garger SJ, Turpen TH, Kumagai MH;
PI	WPI: 2002-681656/73.
DR	N-PADB; AAD45225.
DR	Novel human alpha-galactosidase polypeptide useful for treating
PT	lysosomal storage diseases -
PT	Claim 7; Page 49-50; 88pp; English..
XX	The invention relates to human alpha-galactosidase truncated at the
CC	carboxy terminus and the production of enzymatically active recombinant
CC	human and animal lysosomal enzymes. The invention is useful for producing
CC	lysosomal enzymes for treating lysosomal storage diseases, producing
CC	altered or mutated proteins, enzymatically active or otherwise, to serve
CC	as precursors or substrates for further in vivo or in vitro processing
CC	to a specialised industrial form for research or therapeutic uses, to
CC	produce more effective therapeutic enzyme, for producing antibodies
CC	against lysosomal enzymes for medical diagnostic use, and in any
CC	commercial process that involves substrate hydrolysis. The present
CC	sequence is human rGAL-12 protein.
XX	
SQ	Sequence 417 AA;
	Query Match 98.8%; Score 2293; DB 23; Length 417;
	Best Local Similarity 100.0%; Pred. No. 5.4e-220;
	Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MQLRNPEHLHGCAALRALFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
DB	1 MQLRNPEHLHGCAALRALFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
QY	61 DSCISEKLFWEMAEIMVSEGWKDAGYEYLCTIDCWMAQRDSEGRLOADPQRFPHGIQL 120
DB	61 DSCISEKLFWEMAEIMVSEGWKDAGYEYLCTIDCWMAQRDSEGRLOADPQRFPHGIQL 120
QY	121 ANTVHSKGLKGIYADVGNKTACGFGSGGYDIDAQTFADWGVDLLKFDCYCDSLENL 180
DB	121 ANTVHSKGLKGIYADVGNKTACGFGSGGYDIDAQTFADWGVDLLKFDCYCDSLENL 180
QY	181 ADGYKHMSALNRTGRSIVYSCEWPLYMFPQPNYTEIRQYCNHNWFADIDDWSKSII 240
DB	181 ADGYKHMSALNRTGRSIVYSCEWPLYMFPQPNYTEIRQYCNHNWFADIDDWSKSII 240
QY	241 SILDWTSFNGEIRVDVAGPGGWNDDMLVTGNFGLSNQQVOTOMALWAIAAPLFMSNDL 300
DB	241 SILDWTSFNGEIRVDVAGPGGWNDDMLVTGNFGLSNQQVOTOMALWAIAAPLFMSNDL 300
QY	301 RHISPOKALLQDKDVIAINQDLPGKGQYLRQGDNFVEWERPLSLGLAWAVAMINRQEI 360
DB	301 RHISPOKALLQDKDVIAINQDLPGKGQYLRQGDNFVEWERPLSLGLAWAVAMINRQEI 360
QY	361 GPRSYTIAVASLGKGVACNPACFITOLLPVKRKLGFYEWTSLRSHINPTGTVLILQL 417
DB	361 GPRSYTIAVASLGKGVACNPACFITOLLPVKRKLGFYEWTSLRSHINPTGTVLILQL 417
RESULT 6	
AAE28210	
ID	AAE28210 standard; Protein; 421 AA.
XX	AAE28210;
AC	
XX	27-DEC-2002 (first entry)
DT	
DE	Human rGAL-8 protein.
XX	
XX	Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW	therapeutic; rGAL-8.
XX	
OS	Homo sapiens.

```
xx Key Key Location/Qualifiers
FH Cleavage-site 401..402
FT /note= "Ctpp cleavage site"
XX
XX US2002088024-A1.
XX
XX 04-JUL-2002.
XX
XX 13-NOV-2001; 2001US-099303059.
XX
XX 26-JUL-2000; 2000US-0626127.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2002-681656/73.
XX N-PSDB; AAD45223.
XX
XX Novel human alpha-galactosidase polypeptide useful for treating
XX lysosomal storage diseases
XX
XX Claim 7; Page 44; 88pp; English.
XX
XX The invention relates to human alpha-galactosidase truncated at the
XX carboxy terminus and the production of enzymatically active recombinant
XX human and animal lysosomal enzymes. The invention is useful for producing
XX lysosomal enzymes for treating lysosomal storage diseases, producing
XX altered or mutated proteins, enzymatically active or otherwise, to serve
XX as precursors or substrates for further in vivo or in vitro processing
XX to a specialised industrial form for research or therapeutic uses, to
XX produce more effective therapeutic enzyme, for producing antibodies
XX against lysosomal enzymes for medical diagnostic use, and in any
XX commercial process that involves substrate hydrolysis. The present
XX sequence is human rGAL-8 protein.
XX
XX Sequence 421 AA;
XX
Query Match 98.8%; Score 2293; DB 23; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.5e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPEHLGICALALFLALVSWDIPGARALDNLARTPTMGWLHWEFMCNLDCCQEP 60
Db 1 MOLRNPEHLGICALALFLALVSWDIPGARALDNLARTPTMGWLHWEFMCNLDCCQEP 60
Qy 61 DSCISEKLFEMAEIMVSEGWKDAYEYLCIDDCWMAQPDSEGRQLQADPQRFPHGIRQL 120
Db 61 DSCISEKLFEMAEIMVSEGWKDAYEYLCIDDCWMAQPDSEGRQLQADPQRFPHGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGTGGSGFYDIDAQTFADWGVLDLKFDCGYCDSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGTGGSGFYDIDAQTFADWGVLDLKFDCGYCDSLENL 180
Qy 181 ADGKXKMSALNRTGRSIVISCEWPLYMWPFOKPNYTEIROYCNHNRNFADIDDSWSIK 240
Db 181 ADGKXKMSALNRTGRSIVISCEWPLYMWPFOKPNYTEIROYCNHNRNFADIDDSWSIK 240
Qy 241 SILDWTSEFQERIVDVAGPGWNPDMVLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSEFQERIVDVAGPGWNPDMVLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKQDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKQDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQEIG 360
Qy 361 GPRSYTTAVASLGKGVACNPACFTTOLLPVKRKLGFEYEWTSRLRSHINPTGVLLQL 417
Db 361 GPRSYTTAVASLGKGVACNPACFTTOLLPVKRKLGFEYEWTSRLRSHINPTGVLLQL 417
```

```
RESULT 7
AAR63234
ID AAR63234 standard; Protein; 429 AA.
XX
XX AAR63234;
XX
XX 29-JUN-1995 (first entry)
XX
XX Human alpha-galactosidase A protein.
XX
XX Human; alpha-galactosidase A; expression; mammalian; expression plasmid;
XX chromatography; fusion protein; Staphylococcus; protein A; cleavage site;
XX housekeeping gene; collagenase; affinity chromatography; galactosidase;
XX enzyme replacement therapy; gene therapy; lysosomal storage disorder;
XX Fabry disease; alpha-D-galacto-glycoconjugate; blood group.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Modified-site 139..141
FT /note= "N-linked glycosylation"
FT Modified-site 192..194
FT /note= "N-linked glycosylation"
FT Modified-site 215..217
FT /note= "N-linked glycosylation"
FT Modified-site 408..410
FT /note= "N-linked glycosylation"
XX
XX US3356804-A.
XX
XX 18-OCT-1994.
XX
XX 24-OCT-1990; 90US-0602824.
XX
XX 24-OCT-1990; 90US-0602824.
XX (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
XX Bishop DF, Desnick RJ, Ioannou YA;
XX
XX WPI; 1994-340330/42.
XX N-PSDB; AAQ77861.
XX
XX Recombinant human alpha-galactosidase A prodn. - using a
XX mammalian host cell expression system to obtain high yields of
XX enzymatically active enzyme
XX
XX Claim 3; Column 43-46; 60pp; English.
XX
XX The amino acid sequence of the human alpha-galactosidase A (hAGA).
XX The gene encodes a protein of 429 amino acids. Methods to overexpress
XX and purify the protein include: (i) placing the 1.45 kb human hAGA gene
XX fragment in a mammalian expression plasmid e.g. AAP91023 and purifying
XX total cell protein by chromatographic methods including affinity
XX chromatography or (ii) by producing a fusion protein comprising hAGA
XX fused to easily purified protein sequences e.g. the Staphylococcus
XX protein A, with cleavage site for a "housekeeping" gene e.g. collagenase
XX cleavage site, inserted between the two sequences in a mammalian
XX expression plasmid. The fusion protein can be easily purified by
XX affinity chromatography and the galactosidase protein released from the
XX fusion protein whilst still bound to the column by cleavage with
XX collagenase. The methods result in active alpha-galactosidase A
XX protein. The hAGA can be used for enzyme replacement therapy in patients
XX with the lysosomal storage disorder, Fabry disease. The protein can also
XX be used in vitro to modify alpha-D-galacto-glycoconjugates in a variety
XX of processes e.g. converting blood group B erythrocytes to group O.
XX
XX Sequence 429 AA;
XX
Query Match 98.8%; Score 2293; DB 15; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEEP 60
 DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEEP 60
 QY 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSGRGLQADPQRFPHGIRQL 120
 DB 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSGRGLQADPQRFPHGIRQL 120
 QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYDYDIDAOQTFADGWGVDLLKFDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYDYDIDAOQTFADGWGVDLLKFDGCYCDLSLENL 180
 QY 181 ADGKXKMSLALNRTGRSIVYSCWPLMYMPFQKPNYTEIROYNHWNFNADIDDSWKSIX 240
 DB 181 ADGKXKMSLALNRTGRSIVYSCWPLMYMPFQKPNYTEIROYNHWNFNADIDDSWKSIX 240
 QY 241 SILDWTSEFNOERIVDVAGPGWNDPDMVLVGNFGLSNQOQVOMALWAIMAAPLFMSNDL 300
 DB 241 SILDWTSEFNOERIVDVAGPGWNDPDMVLVGNFGLSNQOQVOMALWAIMAAPLFMSNDL 300
 QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFVWERPLSGLAWAVAMINRQIEG 360
 DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFVWERPLSGLAWAVAMINRQIEG 360
 QY 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
 DB 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 8

AA053765
 ID AAR53765 standard; Protein: 429 AA.

AC AAR53765;

DT 25-MAR-2003 (updated)

DT 15-NOV-1994 (first entry)

XX Alpha-galactosidase A.

DE Alpha-galactosidase A; Fabry disease; blood group O; blood group B;
 KW CHO; Chinese hamster ovary; enzyme replacement therapy;
 KW glycoconjugate; alpha-Gal A.

XX Homo sapiens.

Key Location/Qualifiers
 FT Modified-site 236 /note= "putative N-glycosylation site"
 FT Modified-site 305 /note= "putative N-glycosylation site"
 FT Modified-site 312 /note= "putative N-glycosylation site"
 FT Modified-site 345 /note= "putative N-glycosylation site"
 FT Modified-site /note= "putative N-glycosylation site"

XX W09412628-A1.

XX 09-JUN-1994.

XX 30-NOV-1993; 93WO-US11539.

XX 30-NOV-1992; 92US-0983451.

XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX Bishop DF, Desnick RJ, Ioannou YA;

XX WPI; 1994-200257/24.

XX N-PSDB; AA066241.

XX Prodn of human alpha-galactosidase A - by culturing cells contg

PT the coding sequence and the beta-galactosidase alpha-2,6-sialyl
 PT. transferase gene and regulation sequences

XX Disclosure; Page 102-103; 156pp; English.

XX The cDNA sequence (AA066241) of human alpha-galactosidase A
 CC (alpha-Gal A) was determined. The predicted amino acid sequence
 CC (AAR53765) showed 50% homology with human
 CC (alpha-N-acetylgalactosaminidase (AAR53766) (alpha-galactosidase B)
 CC and showed short regions of homology with yeast Mel 1 (AAR53767).
 CC Recombinant alpha-Gal A was produced in CHO cells. The
 CC enzyme can be used for Fabry disease enzyme replacement therapy,
 CC to convert blood group B to O, or to hydrolyze alpha-D-galactosyl
 CC residues from glycoconjugates.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 429 AA;

Query Match 98.8%; Score 2293; DB 15; Length 429;

Best Local Similarity 100.0%; Pred. No. 5.7e-220;

Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEEP 60

DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDQCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSGRGLQADPQRFPHGIRQL 120

DB 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCIDDCWMAPOQDSGRGLQADPQRFPHGIRQL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYDYDIDAOQTFADGWGVDLLKFDGCYCDLSLENL 180

DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYDYDIDAOQTFADGWGVDLLKFDGCYCDLSLENL 180

QY 181 ADGKXKMSLALNRTGRSIVYSCWPLMYMPFQKPNYTEIROYNHWNFNADIDDSWKSIX 240

DB 181 ADGKXKMSLALNRTGRSIVYSCWPLMYMPFQKPNYTEIROYNHWNFNADIDDSWKSIX 240

QY 241 SILDWTSEFNOERIVDVAGPGWNDPDMVLVGNFGLSNQOQVOMALWAIMAAPLFMSNDL 300

DB 241 SILDWTSEFNOERIVDVAGPGWNDPDMVLVGNFGLSNQOQVOMALWAIMAAPLFMSNDL 300

QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFVWERPLSGLAWAVAMINRQIEG 360

DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFVWERPLSGLAWAVAMINRQIEG 360

QY 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

DB 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 9

AA070207

ID AAR70207 standard; Protein: 429 AA.

XX AAR70207;

XX 25-MAR-2003 (updated)

XX 21-SEP-1995 (first entry)

XX Alpha-galactosidase of human placenta.

DE Alpha-galactosidase; coffee bean; CBG; alpha-1,3-linked galactose;

XX B antigen; blood group.

XX Homo sapiens.

XX W09507088-A1.

XX 16-MAR-1995.

XX 26-AUG-1994; 94WO-US09662.

PR 08-SEP-1993; 93US-0118470.
XX (NYBL-) NEW YORK BLOOD CENT INC.
XX Goldstein J, Zhu A;
XX WPT; 1995-123231/16.
XX New recombinant coffee bean alpha-galactosidase - used for
PT cleaving alpha1,3-linked galactose residues on the surface of
PT cells for prodn. of blood prods.
XX Disclosure; Page 45-46; 67pp; English.
XX DNA encoding coffee bean alpha-galactosidase (CBG) was obtained by
XX isolating mRNA from coffee beans, prepq, cDNA and subjecting this
XX to PCR amplification using primers based on the amino acid
XX sequence of CBG. A full-length cDNA clone is given in AAQ83533, which
XX encoded a 42 kDa protein (AA870205). Recombinant CBG was produced in
XX Sf9 cells. CBG showed homology to alpha-galactosidases from guar,
XX human placenta, S. cerevisiae and A. niger (AA870206-09).
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 429 AA;
Query Match 98.8%; Score 2293; DB 16; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHGCALALRFLALVSWDIPGARALONGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPELHGCALALRFLALVSWDIPGARALONGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFEMAEMLVSEGWKADAGYEYLCIDCWMAPORDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFEMAEMLVSEGWKADAGYEYLCIDCWMAPORDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGIYADVGNKTCAGFGPGSGFYDIDAQTFADGWVDLLKFDGICYDLSLENL 180
DB 121 ANYVHSGKLGIYADVGNKTCAGFGPGSGFYDIDAQTFADGWVDLLKFDGICYDLSLENL 180
QY 181 ADGKHKMSALNRTGRSIVYSCWPLYMWPFPQPNYTEIROYNHWRNFADIDDSWSKIK 240
DB 181 ADGKHKMSALNRTGRSIVYSCWPLYMWPFPQPNYTEIROYNHWRNFADIDDSWSKIK 240
QY 241 SILDWTSFNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVITAINODPLGKQGYOLROGDNFVWERPPLSGLAWAVAMINRQEIG 360
DB 301 RHISPOKALLQDKDVITAINODPLGKQGYOLROGDNFVWERPPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
DB 361 GPRSYTTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
RESULT 10.
AAE28206
ID AAE28206 standard; Protein: 429 AA.
XX AAE28206;
XX 27-DEC-2002 (first entry)
XX Human WT rGAL-A protein.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-A.
XX Homo sapiens.
XX

FH Key Location/Qualifiers
FT Cleavage-site 401..402 /note= "CTPP cleavage site"
XX US2002088024-A1.
XX 04-JUL-2002.
XX 13-NOV-2001; 2001US-0993059.
XX 26-JUL-2000; 2000US-0626127.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPT; 2002-681656/73.
XX N-PSDB; AAD45219.
XX Novel human alpha-galactosidase polypeptide useful for treating
XX lysosomal storage diseases
XX Claim 7; Page 31-32; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
XX carboxy terminus and the production of enzymatically active recombinant
XX human and animal lysosomal enzymes. The invention is useful for producing
XX lysosomal enzymes for treating lysosomal storage diseases, producing
XX altered or mutated proteins, enzymatically active or otherwise, to serve
XX as precursors or substrates for further in vivo or in vitro processing
XX to a specialised industrial form for research or therapeutic uses to
XX produce more effective therapeutic enzyme, for producing antibodies
XX against lysosomal enzymes for medical diagnostic use, and in any
XX commercial process that involves substrate hydrolysis. The present
XX sequence is human WT rGAL-A protein.
XX Note: This sequence SEQ ID NO:4 shown in the sequence listing differs
XX from that shown in page 32-33. This sequence contains an additional
XX Met residue at the N-terminus.
XX SQ Sequence 429 AA;
Query Match 98.8%; Score 2293; DB 23; Length 429;
Best Local Similarity 100.0%; Pred. No. 5.7e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFEMAEMLVSEGWKADAGYEYLCIDCWMAPORDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFEMAEMLVSEGWKADAGYEYLCIDCWMAPORDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGIYADVGNKTCAGFGPGSGFYDIDAQTFADGWVDLLKFDGICYDLSLENL 180
DB 121 ANYVHSGKLGIYADVGNKTCAGFGPGSGFYDIDAQTFADGWVDLLKFDGICYDLSLENL 180
QY 181 ADGKHKMSALNRTGRSIVYSCWPLYMWPFPQPNYTEIROYNHWRNFADIDDSWSKIK 240
DB 181 ADGKHKMSALNRTGRSIVYSCWPLYMWPFPQPNYTEIROYNHWRNFADIDDSWSKIK 240
QY 241 SILDWTSFNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVITAINODPLGKQGYOLROGDNFVWERPPLSGLAWAVAMINRQEIG 360
DB 301 RHISPOKALLQDKDVITAINODPLGKQGYOLROGDNFVWERPPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
DB 361 GPRSYTTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 11
AAV48575
ID AAY48575 standard; Protein: 478 AA.
AC AAY48575;
XX
XX
DT 08-DEC-1999 (first entry)
XX
DE Human breast tumour-associated protein 36.
XX
XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW treatment; tumour; cytostatic; medicament.
XX
XX Homo sapiens.
OS
XX
PN DE19813839-A1.
XX
PD 23-SEP-1999.
XX
XX 20-MAR-1998; 98DE-1013839.
PF
XX 20-MAR-1998; 98DE-1013839.
PR
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA
XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
XX WPI: 1999-528981/45.
DR N-PSDB; AAZ33649.
XX
XX Human nucleic acid sequences and protein products from tumor breast
PT tissue, useful for breast cancer therapy -
XX
XX Claim 25; 158-159; 188pp; German.
XX
XX This invention describes novel human nucleic acid sequences from tumor
CC breast tissue which have cytostatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer. AAY48540-Y48617 represent protein
CC fragments encoded by the expressed sequence tags described in the method
CC of the invention.
XX
SQ Sequence 478 AA;

Query Match 98.8%; Score 2293; DB 20; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.6e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 60
Db 50 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 109
QY 61 DSCISEKLEFMEAEMLMVSEGKWDAGYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIROL 120
Db 110 DSCISEKLEFMEAEMLMVSEGKWDAGYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIROL 169
QY 121 ANYVHSKGLKGLIYADVGNKTCAGPFGSFGYDYIDAQTFADWGVLLKFDGVCYCDSENL 180
Db 170 ANYVHSKGLKGLIYADVGNKTCAGPFGSFGYDYIDAQTFADWGVLLKFDGVCYCDSENL 229
QY 181 ADGYKHMSALNRTGRSIVYSCWPLYMFPKPNTEIRQYCNHWRNFADIDDSWKSITK 240
Db 230 ADGYKHMSALNRTGRSIVYSCWPLYMFPKPNTEIRQYCNHWRNFADIDDSWKSITK 289
QY 241 SILDWTSFNOERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 290 SILDWTSFNOERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFMSNDL 349

QY 301 RHISPOAKALLQDKDVIAINQDPLGKQYQLRQGNFVWERPLSLGLAVAMINRQETG 360
Db 350 RHISPOAKALLQDKDVIAINQDPLGKQYQLRQGNFVWERPLSLGLAVAMINRQETG 409
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
Db 410 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 466

RESULT 12
AAR07305
ID AAR07305 standard; protein: 429 AA.
XX
AC AAR07305;
XX
XX 25-MAR-2003 (updated)
DT 29-JAN-1991 (first entry)
XX
DE Precursor form of alpha-galactosidase A.
XX
XX Baculovirus expression vectors; human alpha-galactosidase A;
KW Fabry disease.
XX
XX Autographa californica.
XX
XX WO9011353-A.
XX
XX 04-OCT-1990.
XX
XX 23-MAR-1990; 90WO-US01571.
XX
XX 24-MAR-1989; 89US-0328421.
XX
XX (RESE) RESEARCH CORP TECHNOLOGIES INC.
PA
XX Calhoun DH, Coppola G;
XX
XX WPI: 1990-320257/42.
DR N-PSDB; AAQ06165.
XX
XX Baculovirus expression vectors - for prodn. of human alpha-
PT galactosidase A, useful for treatment of Fabry disease
XX
XX Claim 4; page 30; 41pp; English.
XX
XX This precursor protein of human alpha-galactosidase is useful in the
CC treatment of diseases associated with deficiency of alpha-galacto-
CC sidase A in mammals, e.g. Fabry disease. Dosage is 10-1000 microg.
CC Monoclonal antibodies raised against the precursor are useful for
CC detecting such diseases. It is produced in large amts. using
CC recombinant DNA methods.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 429 AA;

Query Match 98.6%; Score 2290; DB 11; Length 429;
Best Local Similarity 99.8%; Pred. No. 1.1e-219;
Matches 416; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFHFMCLDCQEEP 60
QY 61 DSCISEKLEFMEAEMLMVSEGKWDAGYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIROL 120
Db 61 DSCISEKLEFMEAEMLMVSEGKWDAGYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIROL 170
QY 121 ANYVHSKGLKGLIYADVGNKTCAGPFGSFGYDYIDAQTFADWGVLLKFDGVCYCDSENL 180
Db 121 ANYVHSKGLKGLIYADVGNKTCAGPFGSFGYDYIDAQTFADWGVLLKFDGVCYCDSENL 180
QY 181 ADGYKHMSALNRTGRSIVYSCWPLYMFPKPNTEIRQYCNHWRNFADIDDSWKSITK 240

Db 181 ADGKXMSLALNRTGRSIVSYSCWPLVYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTFSFNGERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIWAAPLFMSNDL 300
Db 241 SILDWTFSFNGERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIWAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINODPLGKQGYOLROGDNFVWWRPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYOLROGDNFVWWRPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 13

AAE28208
ID AAE28208 standard; Protein: 424 AA.

XX AC AAE28208;

XX DT 27-DEC-2002 (first entry)

XX DE Human rGAL-4 protein.

XX KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX KW therapeutic; rGAL-4.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT Cleavage-site 400..401
XX FT /note= "CTPP cleavage site"

XX PN US2002088024-A1.

XX PD 04-JUL-2002.

XX PF 13-NOV-2001; 2001US-0993059.

XX PR 26-JUL-2000; 2000US-0626127.

XX PA (GARG/) GARGER S J.

XX PA (TURP/) TURPEN T H.

XX PA (KUMA/) KUMAGAI M H.

XX PI Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2002-681656/73.

XX DR N-PSDB; AAD45221.

XX PT Novel human alpha-galactosidase polypeptide useful for treating
XX PT lysosomal storage diseases

XX PS Claim 7; Page 36-38; 88pp; English.

XX CC The invention relates to human alpha-galactosidase truncated at the
XX CC carboxy terminus and the production of enzymatically active recombinant
XX CC human and animal lysosomal enzymes. The invention is useful for producing
XX CC lysosomal enzymes for treating lysosomal storage diseases, producing
XX CC altered or mutated proteins, enzymatically active or otherwise, to serve
XX CC as precursors or substrates for further in vivo or in vitro processing
XX CC to a specialised industrial form for research or therapeutic uses, to
XX CC produce more effective therapeutic enzyme, for producing antibodies
XX CC against lysosomal enzymes for medical diagnostic use, and in any
XX CC commercial process that involves substrate hydrolysis. The present
XX CC sequence is human rGAL-4 protein.

XX CC Note: This sequence SEQ ID NO:8 shown in the sequence listing differs
XX CC from that shown in page 38-39. This sequence contains an additional
XX CC Met residue at the N-terminus.

XX SQ Sequence 424 AA;

Query Match 98.5%; Score 2288; DB 23; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.8e-219;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 QLRNPELHLCALALREFALALVSWDIPGARALDNLGLARTPTMGWLHWFERNLDCQEPD 61
Db 1 QLRNPELHLCALALREFALALVSWDIPGARALDNLGLARTPTMGWLHWFERNLDCQEPD 60
Qy 62 SCISEKLFMEAEMLWSEGKWDAGYEYLCIDDCWMAFQDSRQADPQRFPHGIROLA 121
Db 61 SCISEKLFMEAEMLWSEGKWDAGYEYLCIDDCWMAFQDSRQADPQRFPHGIROLA 120
Qy 122 NYHSGKLGIGIYADVGNKTCAGPPGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSLELA 181
Db 121 NYHSGKLGIGIYADVGNKTCAGPPGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSLELA 180
Qy 182 DGYKXMSLALNRTGRSIVSYSCWPLVYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 241
Db 181 DGYKXMSLALNRTGRSIVSYSCWPLVYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 242 ILDTWTFNQERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIWAAPLFMSNDL 301
Db 241 ILDTWTFNQERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIWAAPLFMSNDL 300
Qy 302 HISPOAKALLQDKDVIAINODPLGKQGYOLROGDNFVWWRPLSLGLAWAVAMINRQEI 361
Db 301 HISPOAKALLQDKDVIAINODPLGKQGYOLROGDNFVWWRPLSLGLAWAVAMINRQEI 360
Qy 362 PRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 PRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 416

RESULT 14

AAW40801
ID AAW40801 standard; peptide: 429 AA.

XX AC AAW40801;

XX DT 31-MAR-1998 (first entry)

XX DE Alpha-galactosidase A.

XX KW Alpha-galactosidase A; human; Fabry disease; X-linked deficiency;
XX KW gene therapy; glycosylated protein; leader sequence.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT Peptide 1..31
XX FT /note= "leader sequence"
XX FT Protein 32..429

XX PN US5658567-A.

XX PD 19-AUG-1997.

XX PF 22-NOV-1993; 93US-0155734.

XX PR 07-MAY-1990; 90US-0520312.

XX PR 24-MAR-1989; 89US-0328421.

XX PR 28-SEP-1992; 92US-0951594.

XX PR 22-NOV-1993; 93US-0155734.

XX PA (RESE) RESEARCH CORP TECHNOLOGIES INC.

XX PI Calhoun DH, Coppola G;

XX DR WPI; 1997-424181/39.

XX PT Recombinant, biologically active human alpha-galactosidase produced
XX PT in baculovirus infected cells - at high level and in glycosylated

form, used for treatment of Fabry disease

Claim 3; Column 19-20; 14pp; English.

This sequence represents the human alpha-galactosidase A. The leader sequence of this protein can be replaced with the sequence shown in AAW40800. Recombinant glycosylated, biologically active human alpha-galactosidase A is produced in a method of the invention by culturing host cells containing recombinant baculovirus that includes a nucleic acid encoding the precursor or mature form of alpha-galactosidase A. This sequence is used to treat Fabry disease, an X-linked deficiency of alpha-galactosidase A. Also contemplated is gene therapy using the nucleic acid sequence encoding it. A baculovirus system provides high level, stable expression of the protein, most of which is secreted into the medium. This protein is produced with greater activity than in *E. coli* (probably because it is glycosylated).

Query Match 98.3%; Score 2282; DB 18; Length 429;
Best Local Similarity 99.5%; Pred. No. 7.1e-219;
Matches 415; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCQEEP 60
Db 1 MGLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCQEEP 60
QY 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFGCYCDSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFGCYCDSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIATINQDPLGKQGYLROGDNFVWPERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYLROGDNFVWPERPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 15
AAR31898
ID AAR31898 standard; Protein: 428 AA.
XX
AC AAR31898;
DT
DT 25-MAR-2003 (updated)
DT 01-JUN-1993 (first entry)
XX
DE Human alpha-galactosidase A.
XX
XX Lysosomal storage disease; globotriaosylceramide; glycolipid;
KW hydrolysis; X-chromosome; X-linked deficiency;
KW enzyme replacement therapy.
XX
XX Homo sapiens.
XX
XX US5179023-A.
XX
XX 12-JAN-1993.

XX PF 07-MAY-1990; 90US-0520312.
XX PR 24-MAR-1989; 89US-0328421.
XX PR 07-MAY-1990; 90US-0520312.
XX PA (RESE) RESEARCH CORP TECHNOLOGIES INC.
XX PI Calhoun DH, Coppola G;
XX DR WPI; 1993-044850/05.
XX DR N-PSDB; AAO36372.
XX PT DNA vector encoding precursor alpha-galactosidase A - useful for
XX PT prodn. of recombinant enzyme, for treating Fabry disease
XX PS Claim 3; Column 19; 20pp; English.
XX
XX The human alpha-galactosidase coding region is contained in clone
CC M13.L21.1 (Hantzopoulos et al., Gene 57:159-169 (1987)). The
CC invention covers expression vectors comprising DNA encoding a
CC precursor form of alpha-galactosidase A (pref. having the amino
CC acid sequence AAR31898) operably linked to a baculovirus sequence.
CC Recombinantly produced alpha-galactosidase A will be
CC useful for enzyme replacement therapy of Fabry disease. The disease
CC is caused by a deficiency of alpha-galactosidase A which results in
CC an accumulation of globotriaosylceramide (the major sphingolipid
CC substrate of the enzyme) and related glycolipids. Deposition of the
CC substrate leads to ischemia and infarction and early demise due to
CC vascular disease of the heart and kidney. The nucleotide sequence
CC printed in the specification and stated to encode alpha-
CC galactosidase A contains a number of apparent anomalies (see
CC AAO36372); the amino acid sequence AAR31898 is identical to that printed
CC in the specification and not that directly decoded from AAO36372.
XX
XX Sequence 428 AA;
Query Match 98.0%; Score 2275; DB 14; Length 428;
Best Local Similarity 99.3%; Pred. No. 3.5e-218;
Matches 414; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCQEEP 60
QY 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFGCYCDSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFGCYCDSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIATINQDPLGKQGYLROGDNFVWPERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYLROGDNFVWPERPLSGLAWAVAMINRQEIG 360
QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

Search completed: July 29, 2003, 16:14:24
Job time : 116.503 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:12:30 ; Search time 47.3154 Seconds
(without alignments)
859.748 Million cell updates/sec

Title: US-09-993-059-18
Perfect score: 2322
Sequence: 1 MOLNPELHGCALALRFLA.....RSHNPVTGTVLLQSEKDEL 423
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	98.8	429	1 GBHUA	alpha-galactosidas
2	1806	77.8	419	2 JC4522	alpha-galactosidas
3	1054	45.4	405	2 S45522	alpha-N-acetylglala
4	952.5	41.0	411	2 A33485	alpha-N-acetylglala
5	930	40.1	358	2 A33265	hypothetical prote
6	861	37.1	451	2 T24018	alpha-galactosidas
7	716.5	30.9	434	2 T47748	alpha-galactosidas
8	712	30.7	378	2 T50781	alpha-galactosidas
9	687.5	29.6	422	2 T06388	alpha-galactosidas
10	682	29.4	435	2 T10860	alpha-galactosidas
11	681.5	29.3	411	2 S07472	alpha-galactosidas
12	676	29.1	680	2 T36472	probable secreted
13	609	26.2	436	2 T39118	probable alpha-gal
14	598	25.8	396	2 JC5558	alpha-galactosidas
15	557	24.0	469	2 S45453	alpha-galactosidas
16	543.5	23.4	545	2 S23582	alpha-galactosidas
17	521	22.4	471	2 S50312	alpha-galactosidas
18	519	22.4	471	2 S50311	alpha-galactosidas
19	516	22.2	444	2 S74221	alpha-galactosidas
20	514	22.1	471	2 J01021	alpha-galactosidas
21	510	22.0	471	2 S50310	alpha-galactosidas
22	502	21.6	471	1 GBBVAG	alpha-galactosidas
23	300.5	12.9	204	2 T04423	probable alpha-gal
24	282.5	12.2	432	2 F83883	hypothetical prote
25	198	8.5	159	2 T04422	alpha-galactosidas
26	180.5	7.8	624	2 S74222	alpha-galactosidas
27	129	5.6	641	2 A55349	glucan 1,6-alpha-1
28	125	5.4	348	2 A70311	hypothetical prote
29	116	5.0	4199	2 S76412	hypothetical prote

ALIGNMENTS

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N;Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1987 #sequence.revision 27-Oct-1995 #text.change 18-Jun-1999
C;Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R;Kornreich, R.; Desnick, R.J.; Bishop, D.F.
Nucleic Acids Res. 17, 3301-3302, 1989
A;Title: Nucleotide sequence of the human alpha-galactosidase A gene.
A;Reference number: S04081; MUID:89263745; PMID:2542896
A;Accession: S04081
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-429 <KOR>
A;Cross-references: EMBL:X14448; NID:g31755; PIDN:CAA32617.1; PID:g31756
R;Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.
Gene 58, 177-188, 1987
A;Title: A genomic clone containing the promoter for the gene encoding the human lyso
A;Reference number: A29608; MUID:88112869; PMID:2892762
A;Accession: A29608
A;Molecule type: DNA
A;Residues: 1-64 <QUI>
A;Cross-references: GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:g553299
R;Bishop, D.F.; Kornreich, R.; Desnick, R.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988
A;Title: Structural organization of the human alpha-galactosidase A gene: further evi
A;Reference number: A30214; MUID:88234528; PMID:2836863
A;Accession: A30214
A;Molecule type: DNA
A;Residues: 1-64 <BIS>
A;Cross-references: EMBL:M20317; EMBL:J03249
R;Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.
FEBS Lett. 259, 353-356, 1990
A;Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activit
A;Reference number: S14879; MUID:90092580; PMID:2152885
A;Accession: S14879
A;Molecule type: mRNA
A;Residues: 1, 'K', 3-39, 'S', 41-429 <KOI>
A;Cross-references: EMBL:X16889
A;Experimental source: Fabry's disease patient
R;Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick,
Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986
A;Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding th
A;Reference number: A00896; MUID:86259694; PMID:3014515
A;Accession: A00896
A;Molecule type: mRNA
A;Residues: 27-429 <BI2>
A;Cross-references: GB:M13571; NID:g178245; PIDN:AAA51676.1; PID:g178246
A;Experimental source: lung
A;Accession: B00896
A;Molecule type: protein

alpha-amylase (EC
alpha-galactosidas
Subtilase family p
hypothetical prote
hypothetical prote
alpha-amylase (imp
hypothetical prote
alpha-galactosidas
cellulase - Cellul
transglycosylase,
alpha-amylase (EC
N-acetylglactosam
hypothetical prote
ATP synthase subun
hypothetical prote

A:Residues: 32-55,'S',57-58,'R',60-65,'S',67-68;228-232,'N',234-237,'A';298-326;'L',334-
R:Tsuji, S.; Martin, B.M.; Kaslow, D.C.; Migeon, B.R.; Choudary, P.V.; Stubblied, B.K
Eur. J. Biochem. 165, 275-280, 1987
A:Title: Signal sequence and DNA-mediated expression of human lysosomal alpha-galactosidase
A:Reference number: 137140; MUID:87246603; PMID:3036505
A:Accession: 137140
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-429 <RES>
A:Cross-references: EMBL:X05790; NID:g28535; PIDN:CAA29232.1; PID:g757912
C:Genetics:
A:Gene: GDB:GLA
A:Cross-references: GDB:119272; OMIM:301500
A:Map position: Xq21.3-Xq22
A:Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C:Superfamily: alpha-galactosidase
C:Keywords: Fabry disease; glycolipid metabolism; glycoprotein; glycosidase; hydrolase;
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-429/Product: alpha-galactosidase A #status predicted <MAT>
F:139,192,215,408/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 98.8%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 6.9e-188;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPDLHGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MQLRNPDLHGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DCSISEKLFMEALMVSEGWKADAGEYLCIDCWMAPODSEGRLOADPQRPFGIROL 120
Db 61 DCSISEKLFMEALMVSEGWKADAGEYLCIDCWMAPODSEGRLOADPQRPFGIROL 120
QY 121 ANYVHSGKLGIYADVGNKTCAGFPFGSGYDIDAQTFADGWVLLKFDGCGYCDLSLENI 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFPFGSGYDIDAQTFADGWVLLKFDGCGYCDLSLENI 180
QY 181 ADGKHMALNLTGRSIVYSCWPLYMPFPKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
Db 181 ADGKHMALNLTGRSIVYSCWPLYMPFPKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
QY 241 SILDTSFNOERIVDVAGPGWNDPMLVIGNGLSWNOQVOTOMALWAMAAPLFMSNDL 300
Db 241 SILDTSFNOERIVDVAGPGWNDPMLVIGNGLSWNOQVOTOMALWAMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRGDNFVWRPLSLGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRGDNFVWRPLSLGLAWAVAMINRQETG 360
QY 361 GPRSYYTIAVSLGKGVACNPAFCITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
Db 361 GPRSYYTIAVSLGKGVACNPAFCITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
RESULT 2
JC4522
alpha-galactosidase (EC 3.2.1.22) A precursor - mouse
N:Alternate names: alpha-D-galactoside galactohydrolase
C:Species: Mus musculus (house mouse)
C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 22-Jun-1999
C:Accession: JC4522
R:Ohshima, T.; Murray, G.J.; Nagle, J.W.; Quirk, J.M.; Kraus, M.H.; Barton, N.W.; Brady,
Gene 166, 277-280, 1995
A:Title: Structural organization and expression of the mouse gene encoding alpha-galacto
A:Reference number: JC4522; MUID:96125203; PMID:8543175
A:Accession: JC4522
A:Molecule type: mRNA
A:Residues: 1-419 <OHS>
A:Cross-references: GB:U34071; NID:g1141787; PIDN:AAA96749.1; PID:g1141788
A:Experimental source: kidney, C57BL
C:Comment: This enzyme is a lysosomal enzyme that hydrolyses the alpha-D-galactosyl resi
C:Genetics:
A:Gene: alpha GalA

A:Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C:Superfamily: alpha-galactosidase
C:Keywords: glycoprotein; glycosidase; hydrolase; lysosome
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-419/Product: alpha-galactosidase A #status predicted <MAT>
F:139,192,215,408/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 77.8%; Score 1806; DB 2; Length 419;
Best Local Similarity 78.2%; Pred. No. 2.6e-146;
Matches 326; Conservative 41; Mismatches 50; Indels 0; Gaps 0;
QY 1 MQLRNPDLHGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MQLRNPDLHGLGALALRFLALVSDIPGARALDNGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DCSISEKLFMEALMVSEGWKADAGEYLCIDCWMAPODSEGRLOADPQRPFGIROL 120
Db 61 DCSISEKLFMEALMVSEGWKADAGEYLCIDCWMAPODSEGRLOADPQRPFGIROL 120
QY 121 ANYVHSGKLGIYADVGNKTCAGFPFGSGYDIDAQTFADGWVLLKFDGCGYCDLSLENI 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFPFGSGYDIDAQTFADGWVLLKFDGCGYCDLSLENI 180
QY 181 ADGKHMALNLTGRSIVYSCWPLYMPFPKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
Db 181 ADGKHMALNLTGRSIVYSCWPLYMPFPKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
QY 241 SILDTSFNOERIVDVAGPGWNDPMLVIGNGLSWNOQVOTOMALWAMAAPLFMSNDL 300
Db 241 SILDTSFNOERIVDVAGPGWNDPMLVIGNGLSWNOQVOTOMALWAMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRGDNFVWRPLSLGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRGDNFVWRPLSLGLAWAVAMINRQETG 360
QY 361 GPRSYYTIAVSLGKGVACNPAFCITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
Db 361 GPRSYYTIAVSLGKGVACNPAFCITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
RESULT 3
S45522
alpha-N-acetylglucosaminidase - chicken
C:Species: Gallus gallus (chicken)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 22-Jun-1999
C:Accession: S45522; S43413
R:Davis, M.O.; Hata, J.; Smith, D.; Walker, J.C.
submitted to the EMBL Data Library, December 1993
A:Reference number: S45522
A:Accession: S45522
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-405 <DAV>
A:Cross-references: EMBL:L18754; NID:g435036; PIDN:AAA16614.1; PID:g435037
R:Davis, M.O.; Hata, D.J.; Smith, D.; Walker, J.C.
Biochim. Biophys. Acta 1216, 296-298, 1993
A:Title: Cloning and sequence of a chicken alpha-N-acetylglucosaminidase gene.
A:Reference number: S43413; MUID:94060104; PMID:8241271
A:Accession: S43413
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-161,'A',163-405 <DA2>
A:Cross-references: GB:L18754
C:Superfamily: alpha-galactosidase
Query Match 45.4%; Score 1054; DB 2; Length 405;
Best Local Similarity 52.8%; Pred. No. 4.1e-82;
Matches 207; Conservative 58; Mismatches 119; Indels 8; Gaps 6;
QY 32 LDNGLARTPTMGWLHWFRCMCLDCQEPDSCISEKLFMEALMVSEGWKADAGEYLCI 91
Db 1 LENGARTPPMGWLAWERFRCNVNCRDPROCISEMFLFMEMADRIAEDGWRELGYKINI 60

QY 92 DDCWAPORDSEGRLOADQPPHGIROLANVYHSGKLGKGIYADVGNKTCAGPFG-SFG 150
Db 61 DDCWAAQORDAEGRLVDPPEPRGRIKALADYVHARGLKLDIYDGLRLTCGGYPGTLLD 120
QY 151 YYDDAOTFADMGVDLLKFDGCDYCDLENLADGKHKSLNRTGRSTIVYSCWPLYMWP 210
Db 121 RVEDAOTFADMGVDLLKFDGCDYSSGKEQ-AQGTPOMARALNSTRGRIVYSCWPLYOGG 179
QY 211 F-QKPNYTEIROYNHNRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDMLV 269
Db 180 LPPRVNYTLGGEICNLNRNYDDIQDSWDSVLSIVDFNODVLPFAGPGHNDPDMLI 239
QY 270 IGNFGLSNQOVTOMALWAIMAAPLFMSNDLRHISPOKALLQDKVTAINDPLGKOGY 329
Db 240 IGNFGLSYEQSRQALMTWTINAPLLMSTDLRTISPAKLLQRLMIQINQDPLGIQGR 299
QY 330 Q-LRQDNFEYWERPLSGLAWAVAMINROETGGPRSYTIAVASLGKGVACNAPACFIQTLL 388
Db 300 RIIEGSHIEVFLPLSQASALVFFSRR-TDMPFRYTTSLAKLGFPMG---AAEYQDV 355
QY 389 PVKRLGFGYEWTSRLRSHINPTGTVLLQLSEK 420
Db 356 YSGKIISGLKTGDNFTVIINSGVVMVYLCPK 387
RESULT 4
A35485
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3; 367/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-411/Product: alpha-N-acetylgalactosaminidase, splice form 2 #status predicted <M
F:124,177,201,359,385,391/Binding site: carbohydrate (Asn) (covalent) #status predict
Query Match 41.0%; Score 952.5; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 1.9e-73;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;
QY 32 LDNGLARTPTMGWLHWRFCMLDCQEEPDPSCISEKLEFMEAEALMVSEGWKDGAYEYLCI 91
Db 18 LDNGLQTPPPGWLAWERFCNINDEDEPKNCISQLEFMEADRAQDGRDMGYTLNI 77
QY 92 DDCWAPORDSEGRLOADQPPHGIROLANVYHSGKLGKGIYADVGNKTCAGPFG-SFG 150
Db 78 DDCWIG-GRDASGRMLPDPKRPFGIPFLADYVHSLGKLGKGIYADMGNTCMGYPGTLLD 136
QY 151 YYDDAOTFADMGVDLLKFDGCDYCDLENLADGKHKSLNRTGRSTIVYSCWPLYMWP 210
Db 137 KVQDAQTFAEWKVDMLKDGCF-STPEERAGQYPKMAAALNATGRPTAFSCSWPAYEGG-195
QY 211 F-QKPNYTEIROYNHNRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDMLV 269
Db 196 LPPRVNYTLGGEICNLNRNYDDIQDSWDSVLSILNFWVEHQDILQPVAGPGHNDPDMLL 255
QY 270 IGNFGLSNQOVTOMALWAIMAAPLFMSNDLRHISPOKALLQDKVTAINDPLGKOGY 329
Db 256 IGNFGLSYEQSRQALMTWTINAPLLMSTDLRTISPAKLLQRLMIQINQDPLGIQGR 315
QY 330 Q-LRQDNFEYWERPLSGLAWAVAMIN-ROETGGPRSYTIAVASLGKGVACNAPACFIQTLL 387
Db 316 RIIEKKSIEVYMRPLSNKASALVFFSRTDM---PYRHSSIGQLN-----FTGVS 364
QY 388 LPVKKRLGFGYEWTSRLRSH-----INPTGTVLLQL 417
Db 365 IYEAQDVYSGDIISGLRDETNETFTVIINSGVVMVYL 400
RESULT 5
A33265
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-358/Product: alpha-N-acetylgalactosaminidase, splice form 1 #status predicted <M
F:124,177,201/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 40.1%; Score 930; DB 2; Length 358;
Best Local Similarity 57.7%; Pred. No. 1.3e-71;
Matches 176; Conservative 40; Mismatches 85; Indels 4; Gaps 4;

QY 92 DDCWAPORDSEGRLOADQPPHGIROLANVYHSGKLGKGIYADVGNKTCAGPFG-SFG 150
Db 61 DDCWAAQORDAEGRLVDPPEPRGRIKALADYVHARGLKLDIYDGLRLTCGGYPGTLLD 120
QY 151 YYDDAOTFADMGVDLLKFDGCDYCDLENLADGKHKSLNRTGRSTIVYSCWPLYMWP 210
Db 121 RVEDAOTFADMGVDLLKFDGCDYSSGKEQ-AQGTPOMARALNSTRGRIVYSCWPLYOGG 179
QY 211 F-QKPNYTEIROYNHNRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDMLV 269
Db 180 LPPRVNYTLGGEICNLNRNYDDIQDSWDSVLSIVDFNODVLPFAGPGHNDPDMLI 239
QY 270 IGNFGLSNQOVTOMALWAIMAAPLFMSNDLRHISPOKALLQDKVTAINDPLGKOGY 329
Db 240 IGNFGLSYEQSRQALMTWTINAPLLMSTDLRTISPAKLLQRLMIQINQDPLGIQGR 299
QY 330 Q-LRQDNFEYWERPLSGLAWAVAMINROETGGPRSYTIAVASLGKGVACNAPACFIQTLL 388
Db 300 RIIEGSHIEVFLPLSQASALVFFSRR-TDMPFRYTTSLAKLGFPMG---AAEYQDV 355
QY 389 PVKRLGFGYEWTSRLRSHINPTGTVLLQLSEK 420
Db 356 YSGKIISGLKTGDNFTVIINSGVVMVYLCPK 387
RESULT 4
A35485
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3; 367/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-411/Product: alpha-N-acetylgalactosaminidase, splice form 2 #status predicted <M
F:124,177,201,359,385,391/Binding site: carbohydrate (Asn) (covalent) #status predict
Query Match 41.0%; Score 952.5; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 1.9e-73;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;
QY 32 LDNGLARTPTMGWLHWRFCMLDCQEEPDPSCISEKLEFMEAEALMVSEGWKDGAYEYLCI 91
Db 18 LDNGLQTPPPGWLAWERFCNINDEDEPKNCISQLEFMEADRAQDGRDMGYTLNI 77
QY 92 DDCWAPORDSEGRLOADQPPHGIROLANVYHSGKLGKGIYADVGNKTCAGPFG-SFG 150
Db 78 DDCWIG-GRDASGRMLPDPKRPFGIPFLADYVHSLGKLGKGIYADMGNTCMGYPGTLLD 136
QY 151 YYDDAOTFADMGVDLLKFDGCDYCDLENLADGKHKSLNRTGRSTIVYSCWPLYMWP 210
Db 137 KVQDAQTFAEWKVDMLKDGCF-STPEERAGQYPKMAAALNATGRPTAFSCSWPAYEGG-195
QY 211 F-QKPNYTEIROYNHNRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDMLV 269
Db 196 LPPRVNYTLGGEICNLNRNYDDIQDSWDSVLSILNFWVEHQDILQPVAGPGHNDPDMLL 255
QY 270 IGNFGLSNQOVTOMALWAIMAAPLFMSNDLRHISPOKALLQDKVTAINDPLGKOGY 329
Db 256 IGNFGLSYEQSRQALMTWTINAPLLMSTDLRTISPAKLLQRLMIQINQDPLGIQGR 315
QY 330 Q-LRQDNFEYWERPLSGLAWAVAMIN-ROETGGPRSYTIAVASLGKGVACNAPACFIQTLL 387
Db 316 RIIEKKSIEVYMRPLSNKASALVFFSRTDM---PYRHSSIGQLN-----FTGVS 364
QY 388 LPVKKRLGFGYEWTSRLRSH-----INPTGTVLLQL 417
Db 365 IYEAQDVYSGDIISGLRDETNETFTVIINSGVVMVYL 400
RESULT 5
A33265
A:Cross-references: GDB:119445; OMIM:104170
A:Map position: 22q11-22q11
A:Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3
C:Superfamily: alpha-galactosidase
C:Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-358/Product: alpha-N-acetylgalactosaminidase, splice form 1 #status predicted <M
F:124,177,201/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 40.1%; Score 930; DB 2; Length 358;
Best Local Similarity 57.7%; Pred. No. 1.3e-71;
Matches 176; Conservative 40; Mismatches 85; Indels 4; Gaps 4;

QY 32 LDNGLARTPTMGWLHWRERMCNLDCEPDSCISEKLFEMAEIMVSEGWKDAGYEYLICI 91
Db 18 LDNGLLQTPMGWLAWERFCNINCDDPKNISEQLFMEMADRMADQGWDRMGYTYLNI 77
QY 92 DDCWMAQRSEGLQADQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGFPG-SFG 150
Db 78 DDCWIG-GRDASGRMPDKRFRPHGIPFLADYVHSLGKLGKIYADMGNFTCMGYPGTLD 136
QY 151 YYDIDAQTFADGWGVDLLKFGDCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLYMWP 210
Db 137 KVWDAQTFAEWKVDMLKLDGCF-STPEEAQGYPRMAAGLNATGRFIAFSCSWPAYEGG 195
QY 211 F-QKPNYTEIROYNHWRNEADIDDSWKSILDTSFNOERVDVAGPGWNPDMVLV 269
Db 196 LPPRVNYSLOADICNLWRNDDIQDSWRSVLSILNWFVEHQDILQPVAGGHNDPMDLL 255
QY 270 IGNFGLSNQOVTQMALWMAAPLPMNSDLRHISPOAKALQDKOVIAINQDPLGKQY 329
Db 256 IGNFGLSLEQSRQAQMALWTVLAAPLLMSTDLRTISAQNMIDILQPLMIKINQDPLGIQGR 315
QY 330 QLROG 334
Db 316 RIHKG 320

RESULT 6
T24018
hypothetical protein R07B7.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T24018
R:Harris, B.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19830
A:Accession: T24018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-451 <WTL>
A:Cross-references: EMBL:Z75955; PIDN:CA800120.1; GSPDB:GN00023; CESP:R07B7.11
A:Experimental source: clone R07B7
C:Genetics:
A:Gene: CESP:R07B7.11
A:Map position: 5
A:Introns: 61/1; 107/3; 317/3; 351/3
C:Superfamily: alpha-galactosidase

Query Match 37.13; Score 861; DB 2; Length 451;
Best Local Similarity 43.8%; Pred. No. 1.4e-65;
Matches 174; Conservative 61; Mismatches 148; Indels 14; Gaps 6;

QY 28 GARALDNLGARTPTMGWLHWRERMCNLDCEPDSCISEKLFEMAEIMVSEGWKDAGYE 87
Db 12 GAFCLDNLGRTPTMGWMSWTFACEIDCVKHPTGCGINEQLYKMDQLVSGGYDKYGYK 71
QY 88 YLCTDDCWMAQRSEGLQADQRFPHGIRQLANYVHSGKLGKIYADVGNKTCAGFPG 147
Db 72 SVHIDDCWSEMERDShGILLVANKTRFPSPGMKALAKYMHDRGLKFGIYEDYGTCTCGYPG 131
QY 148 SFGYDIDAQTFADGWGVDLLKFGDCYCDLSLENLADGKHKMSLALNRTGRSIVYSCWPLY 207
Db 132 SYRHEKVDQTFADWDVYLUKLDGNCIDQ-AMMPIGYPLFEKELNETGRPIIMYSCSWPAY 190
QY 208 MWPFQKP--NYTEIROYNHWRNEADIDDSWKSILDTSFNOERVDVAGPGWNP 264
Db 191 L-IDHPELYNYNLIGKYCNTRWRNFDDINSSWKSIIISYYDKMQDKHIPHCGPGKWH 248
QY 265 PDMVLGNFGLSNQOVTQMALWMAAPLPMNSDLRHISPOAKALQDKOVIAINQDPL 324
Db 249 PDMVLGNKGTIDMSISQFTVWCVWSAPLIMSNDLRIIGDSFKDVLKNEAKINQDPL 308
QY 325 GKQGYQLROGDNFVEWER--PLSG--LAWAVAMINQETGGPRSYTIAVASLKGKGVACN 379
Db 309 GIMGLRNSTDICVYKQTPSPGDKKSFAYLNRNEKEGYKRIEIQLASIG---LTD 365

QY 380 PACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQ 416
Db 366 PAGIYYVDIWSHVDLGLLRSGLSDSIVVSIAPAGSVFFR 402

RESULT 7
T47748
alpha-galactosidase-like protein - Arabidopsis thaliana
N:Alternate names: protein F18021.270
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T47748
R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansoorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24474
A:Accession: T47748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-434 <BEN>
A:Cross-references: EMBL:ATF18021
A:Experimental source: cultivar Columbia; BAC clone F18021
C:Genetics:
A:Map position: 3
A:Introns: 73/2; 93/1; 112/1; 123/3; 158/2; 182/3; 202/2; 224/2; 252/2; 274/1; 301/3;
A:Note: F18021.270
C:Superfamily: alpha-galactosidase

Query Match 30.9%; Score 716.5; DB 2; Length 434;
Best Local Similarity 42.4%; Pred. No. 2.9e-53;
Matches 153; Conservative 54; Mismatches 107; Indels 47; Gaps 9;

QY 11 GCALALRFLALVSWDIPGARALDNLGARTPTMGWLHWRERMCNLDCEPDSCISEKLFEM 70
Db 41 GLVFSKSFNSIYDTSIMGRQLNGLARTPQMGWNSWNFACN-----INETVIK 90
QY 71 EMALMVSEGWKDAGYEYLCIDDCWMAQRSEGLQADQRFPHGIRQLANYVHSGKLGK 130
Db 91 ETADALVSSGLADLGIYHVNIDDCWNSLLRDSGLVPHPETFPFSGIKLLADYVHSGKLGK 150
QY 131 LGIYADVGNKTCAGFPGSGYFYDIDAQTFADGWGVDLLKFGDCYCDLSLENLA---DGYKH 186
Db 151 LGIYSDAGVETCEVHFCSLFEHVDADIFASWGVLYLKYDNCF-----NIGIKPIERYPP 205
QY 187 MSLALNRTGRSIVYS-CEWPL---YMWPFQKPNYTEIROYNHWRNEADIDDSWKSISKI 242
Db 206 MRDALNATGRSIFYSLCEWGVDDPALW-----AKEVNSWRITDDINDTWASMTTI 256
QY 243 LD----WTSFNPQRIVDVAGPGWNPDMVLGNFGLSNQOVTQMALWMAAPLPMNS 298
Db 257 ADLNNKWAAY-----AGPGWNPDMLEIGNGGMTYEEYRGHFSIALMKAPLLIGC 308
QY 299 DLRHISPOAKALQDKOVIAINQDPLGKQGYQLR---QGDNFVEWERPLSLAWAVAMIN 355
Db 309 DVNRMTAETLEILSNKEITIAVNQDPLGVQGRKIQANGENDCCQVWGSGLSDRKNVVALWN 368
QY 356 R 356
Db 369 R 369

RESULT 8
T50781
alpha-galactosidase (EC 3.2.1.22) [imported] - coffee
C:Species: Coffea arabica (coffee)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 18-Aug-2000
C:Accession: T50781
R:Zhu, A.; Goldstein, J.
Gene 140, 227-231, 1994
A:Title: Cloning and functional expression of a cDNA encoding coffee bean alpha-galac-
A:Reference number: Z25235; MUID:94193002; PMID:8144030
A:Accession: T50781
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-378 <2HU>

A:Cross-references: EMBL:L27992; PIDN:AAA33022.1

C:Superfamily: alpha-galactosidase

C:Keywords: glycosidase; hydrolase

Query Match 30.7%; Score 712; DB 2; Length 378;

Best Local Similarity 42.5%; Pred. No. 5.9e-53;

Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;

QY 32 LDNGLARTPTMGWLHWRPNCNLDCEPDSCISEKLFMEMAELMVSEGWKDAYEYLCI 91

Db 16 LANGLGLTPPMGWSNNHFCNLD-----EKLIRETADAMYSTGLAALGYKYNIL 65

QY 92 DDCWAPORSEGRLOADPORPHGIRQLANYVHSGKGLGIYADVGNKTCA-GFPGSFG 150

Db 66 DDCWAEINRDSQNLVPGKSTFSGIKALADYVHSGKGLGIYSDAGTQCTSYMTGSLG 125

QY 151 YYDIDAQTFADMGVDLLKFGDCYCDSENLDADYGHMSLALNRTGRSIVYS-CEWPLYMW 209

Db 126 HEQDAKTFASWGVDYLYKDYCNENNI-SPKERYPTMSKALLNSGRSIFSLCEWG--- 180

QY 210 PFQKPNYETROYCNHWRNFADIDDSWKSISILD-----WTSFNQERIVDVAGPGWNDP 265

Db 181 --BEDPATWAKEYGVNSWRTTGDIDDSWSSMTSRADMDKWASY-----AGPGGWNDP 230

QY 266 DMLVGNFGLSNQOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINODPLG 325

Db 231 DMLVGNNGMTTEYKSHFSIWALAKAPLLIGDIDRSMGATQOLLSNAEVIIVNQDKLG 290

QY 326 KQGYQLRQGDNFVEWPERPLSGLAWAVAMINR-----OEIGGPRSYTTIAVSL 372

Db 291 VQGNKVKYTGDELVWAGPLSGKRVAVALMNRGSSTATITAYWSDVGLPSTAVVNARDL 348

RESULT 9

T06388

alpha-galactosidase (EC 3.2.1.22) - soybean

C:Species: Glycine max (soybean)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Jun-1999

C:Accession: T06388

R:Davis, M.O.; Walker, J.C.; Smith, D.

submitted to the EMBL Data Library, August 1994

A:Description: Cloning and expression of a soybean alpha galactosidase gene.

A:Reference number: Z15645

A:Accession: T06388

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-422 <DAV>

A:Cross-references: EMBL:U12926; NID:g927574; PIDN:AAA73963.1; PID:g927575

A:Experimental source: strain williams

C:Function:

A:Description: catalyzes hydrolysis of melibiose into galactose and glucose

C:Superfamily: alpha-galactosidase

C:Keywords: glycosidase; hydrolase

Query Match 29.6%; Score 687.5; DB 2; Length 422;

Best Local Similarity 43.2%; Pred. No. 8.5e-51;

Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;

QY 32 LDNGLARTPTMGWLHWRPNCNLDCEPDSCISEKLFMEMAELMVSEGWKDAYEYLCI 91

Db 60 LDNGLGLTPPMGWSNNHFCN-----IKEDLIRETADAMYSTGLAALGYQYINI 109

QY 92 DDCWAPORSEGRLOADPORPHGIRQLANYVHSGKGLGIYADVGNKTCA-GFPGSFG 150

Db 110 DDCWAEINRDSQNLVPGKSTFSGIKALADYVHSGKGLGIYSDAGTQCTSYMTGSLG 169

QY 151 YYDIDAQTFADMGVDLLKFGDCYCDSENLDADYGHMSLALNRTGRSIVYS-CEWPLYMW 209

Db 170 HEQDAKTFASWGVDYLYKDYCNENNI-SPKERYPTMSKALLNSGRSIFSLCEWG----- 224

QY 210 PFQKPNYETROYCNHWRNFADIDDSWKSISILD-----WTSFNQERIVDVAGPGWNDP 265

Db

225 --SEDPATWAKSVGNSWRTTGDIDQKWDSMISRADLNDKWASY-----AGPGGWNDP 274

QY 266 DMLVGNFGLSNQOQVOTOMALWAIMAAPLFMSNDLRHISPOAKALLQDKDVIAINODPLG 325

Db 275 DMLVGNNGMTTEYKSHFSIWALAKAPLLIGDIDRSMGATQOLLSNAEVIIVNQDKLG 334

QY 326 KQGYQLRQGDNFVEWPERPLSGLAWAVAMINR 356

Db 335 VQGNKVKYTGDELVWAGPLSGKRVAVALMNR 365

RESULT 10

T10860

alpha-galactosidase (EC 3.2.1.22) - kidney bean

C:Species: Phaseolus vulgaris (kidney bean)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999

C:Accession: T10860

R:Davis, M.O.; Walker, J.C.; Smith, D.

submitted to the EMBL Data Library, August 1994

A:Description: Cloning and expression of a pinto bean alpha galactosidase gene.

A:Reference number: Z17189

A:Accession: T10860

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-425 <DAV>

A:Cross-references: EMBL:U12927; NID:g927576; PID:g927577

C:Function:

A:Description: catalyzes hydrolysis of melibiose into galactose and glucose

C:Superfamily: alpha-galactosidase

C:Keywords: glycosidase; hydrolase

Query Match 29.4%; Score 682; DB 2; Length 425;

Best Local Similarity 41.5%; Pred. No. 2.5e-50;

Matches 144; Conservative 54; Mismatches 113; Indels 36; Gaps 9;

QY 19 LALVSWDIPCARAL-DNGLARTPTMGWLHWRPNCNLDCEPDSCISEKLFMEMAELMV 77

Db 49 MMMSREVDHRRNLVGNGLGQTTPMGWSNNHFCN-----INEDLIRETADAMV 98

QY 78 SEGWDAGYEYLCIDDCWAPORSEGRLOADPORPHGIRQLANYVHSGKGLGIYADV 137

Db 99 STGLAALGYQYINIDCWGELNDSQNLVPGKSTFSGIKALADYVHSGKGLGIYSDA 158

QY 138 GNKTCA-GFPGSFGYDIDAQTFADMGVDLLKFGDCYCDSENLA--DGKHMSLALNRT 194

Db 159 GTQCTSKTTPGSLGHEEQDAKTFASGIDYLYKDYCN-----ENKNISPKERTPPMSKALANS 215

QY 195 GRSIVYS-CEWPLYMWPFQKPNYETROYCNHWRNFADIDDSWKSISILD---WTSFN 249

Db 216 GRPIFFSLCEWG-----SEDPATWAKSVGNSWRTTGDIDQKWDSMISRADLNDWASY- 268

QY 250 QERIVDVAGPGGWNDPDMLVGNFGLSNQOQVOTOMALWAIMAAPLFMSNDLRHISPOAKA 309

Db 269 -----AGPGGWNDPDMLEVGNMGMTTEYKSHFSIWALAKAPLLIGDIDRSMGATQOLLSNAEVIIVNQDKLG 321

QY 310 LLODKQVIAINODPLGKQGYQLRQGDNFVEWPERPLSGLAWAVAMINR 356

Db 322 LLSNEEVIIVNQDKLGVOGKRVKSNNDLELVWAGPLSGKRVAVALMNR 368

RESULT 11

S07472

alpha-galactosidase (EC 3.2.1.22) precursor - guar

C:Species: Cyamopsis tetragonoloba (guar, cluster bean)

C>Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 22-Jun-1999

C:Accession: S07472

R:Overbeek, N.; Fellinger, A.J.; Toonen, M.Y.; van Wassenaar, D.; Verrips, C.T.

Plant Mol. Biol. 13, 541-550, 1989

A:Title: Cloning and nucleotide sequence of the alpha-galactosidase cDNA from Cyamops

A:Reference number: S07472; MUID:91370836; PMID:2577496

A:Accession: S07472

A:Molecule type: mRNA


```

RESULT 15
S45433
alpha-galactosidase (EC 3.2.1.22) MEL precursor - yeast (zygosaccharomyces cidrli)
N:Alternate names: MEL protein
C:Species: Zygosaccharomyces cidrli
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 22-Jun-1999
C:Accession: S45453
R:Turkainen, H.; Hankaapaae, M.; Korhola, M.; Aho, S.
Yeast 10, 733-745, 1994
A:Title: Characterization of MEL genes in the genus Zygosaccharomyces.
A:Reference number: S45453; MUID:95066377; PMID:7975892
A:Accession: S45453
A:Molecule type: DNA
A:Residues: 1-469 <TUR>
A:Cross-references: EMBL:L24957; NID:G538518; PID:g407012
C:Genetics:
A:Gene: MEL
C:Superfamily: alpha-galactosidase
C:Keywords: glycoprotein; glycosidase; hydrolase
F:1-17/Domain: signal sequence #status predicted <SIG>

```

Search completed: July 29, 2003, 16:18:55
Job time : 49.3154 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:17:20 ; Search time 64.349 Seconds
(without alignments)

780.671 Million cell updates/sec

Title: US-09-993-059-18

Perfect score: 2322

Sequence: 1 MQLRNPGLHGCALALRFLA.....RSHNPVTGTVLLQSEKDEL 423

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2322	100.0	423	15	US-10-103-327-18
2	2310	99.5	427	15	US-10-103-327-14
3	2308	99.4	431	15	US-10-103-327-10
4	2306	99.3	435	15	US-10-103-327-6
5	2293	98.8	417	15	US-10-103-327-16
6	2293	98.8	421	15	US-10-103-327-12
7	2288	98.5	424	15	US-10-103-327-8
8	2288	98.5	428	15	US-10-103-327-4
9	2234.5	96.2	409	15	US-10-103-327-22
10	2213	95.3	401	15	US-10-103-327-20
11	2131	91.8	398	15	US-10-165-060-4
12	2131	91.8	398	15	US-10-165-968-4
13	683.5	29.4	682	15	US-10-156-761-9014
14	676.5	29.1	536	15	US-10-156-761-8170
15	430	18.5	658	15	US-10-156-761-9724

16	422	18.2	608	15	US-10-156-761-8616	Sequence 8616, Ap
17	118	5.1	24	15	US-10-103-327-30	Sequence 30, Appl
18	115.5	5.0	700	15	US-10-156-761-9180	Sequence 9180, Ap
19	104	4.5	32	15	US-10-103-327-26	Sequence 26, Appl
20	102	4.4	36	15	US-10-103-327-24	Sequence 24, Appl
21	101.5	4.4	710	15	US-10-156-761-8621	Sequence 8621, Ap
22	100.5	4.3	737	10	US-09-955-732-13	Sequence 13, Appl
23	99.5	4.3	430	11	US-09-917-384-10	Sequence 10, Appl
24	99.5	4.3	430	11	US-09-917-383-10	Sequence 10, Appl
25	95.5	4.1	29	15	US-10-103-327-28	Sequence 28, Appl
26	94.5	4.1	524	10	US-09-761-640-9	Sequence 9, Appl
27	93.5	4.0	436	15	US-10-081-872-16	Sequence 16, Appl
28	93.5	4.0	436	15	US-10-081-872-18	Sequence 18, Appl
29	92	4.0	484	15	US-10-081-872-315	Sequence 315, Ap
30	89.5	3.9	436	15	US-10-081-872-24	Sequence 24, Appl
31	89	3.8	18	15	US-10-103-327-29	Sequence 29, Appl
32	89	3.8	22	15	US-10-103-327-27	Sequence 27, Appl
33	89	3.8	26	15	US-10-103-327-25	Sequence 25, Appl
34	89	3.8	30	15	US-10-103-327-23	Sequence 23, Appl
35	88.5	3.8	626	10	US-09-213-888-21	Sequence 21, Appl
36	88.5	3.8	626	10	US-09-328-877A-21	Sequence 21, Appl
37	87.5	3.8	436	15	US-10-081-872-2	Sequence 2, Appl
38	87.5	3.8	491	10	US-09-881-752A-128	Sequence 128, App
39	87.5	3.8	507	15	US-10-081-872-170	Sequence 170, App
40	87.5	3.8	943	9	US-09-969-362-5	Sequence 5, Appl
41	87.5	3.8	1247	9	US-09-815-242-10145	Sequence 10145, A
42	85.5	3.7	436	15	US-10-081-872-14	Sequence 14, Appl
43	85.5	3.7	436	15	US-10-081-872-44	Sequence 44, Appl
44	85.5	3.7	507	15	US-10-081-872-178	Sequence 178, App
45	84.5	3.6	361	9	US-09-815-242-11251	Sequence 11251, A

ALIGNMENTS

RESULT 1
US-10-103-327-18
; Sequence 18, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match.	100.0%	Score 2322;	DB 15;	Length 423;
Best Local Similarity	100.0%;	Pred. No. 2.5e-231;		
Matches 423;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	MQLRNPGLHGCALALRFLA	VS	WDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEP 60
Db	1	MQLRNPGLHGCALALRFLA	VS	WDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEP 60
Qy	61	DSCTSEKLFMEAEIMVSEGWK	DAGYEYLC	IDDCWMAFQDSEGLQADPQRFPHGIROL 120
Db	61	DSCTSEKLFMEAEIMVSEGWK	DAGYEYLC	IDDCWMAFQDSEGLQADPQRFPHGIROL 120
Qy	121	ANYVHSKGLKGIYADVGNKTC	CAGPGSGFYD	IDAQTFADGWVLLKFDGCYCDLENL 180
Db	121	ANYVHSKGLKGIYADVGNKTC	CAGPGSGFYD	IDAQTFADGWVLLKFDGCYCDLENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
QY 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQSEK 420
QY 421 DEL 423
Db 421 DEL 423

RESULT 2

US-10-103-327-14

; Sequence 14, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO. 14

; LENGTH: 427

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-103-327-14

Query Match 99.58; Score 2310; DB 15; Length 427;
Best Local Similarity 99.1%; Pred. No. 4.4e-230;
Matches 423; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDQCEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDQCEEP 60
QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAQPDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAQPDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
QY 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ --- 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 418 -SEKDEL 423
Db 421 MSEKDEL 427

RESULT 3

US-10-103-327-10

; Sequence 10, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO. 10

; LENGTH: 431

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-103-327-10

Query Match 99.4%; Score 2308; DB 15; Length 431;
Best Local Similarity 98.1%; Pred. No. 7.2e-230;
Matches 423; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDQCEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLARTPTMGWLHWRFCNLDQCEEP 60
QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAQPDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAQPDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
QY 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTOMALWMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNFVWYERPLSGLAWAVAMINRQEI 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ --- 416
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 417 ---LSEKDEL 423
Db 421 MQMSLSEKDEL 431

RESULT 4

US-10-103-327-6

; Sequence 6, Application US/10103327

; Publication No. US20030106095A1

```
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-6

Query Match      99.3%; Score 2306; DB 15; Length 435;
Best Local Similarity 97.2%; Pred. No. 1.2e-229;
Matches 423; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

Qy 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Qy 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSEFNOERIVDVAGPGGWNPDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNPDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIATINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDKDVIATINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTTAVASLGLGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTTAVASLGLGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 6
US-10-103-327-12
; Sequence 12, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match      98.8%; Score 2293; DB 15; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.5e-228;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120

; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match      98.8%; Score 2293; DB 15; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.4e-228;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYGYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHKSLALNRTGRSIVISCEWPLYMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSEFNOERIVDVAGPGGWNPDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNPDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIATINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDKDVIATINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTTAVASLGLGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTTAVASLGLGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 5
US-10-103-327-16
; Sequence 16, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-15

Query Match      98.8%; Score 2293; DB 15; Length 421;
Best Local Similarity 100.0%; Pred. No. 2.5e-228;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHLCALALREFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDDCWMAPOQSEGRLOADPQRFPHGIRQL 120
```

Qy 121 ANVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 180
Db 121 ANVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 180
Qy 181 ADGKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Db 181 ADGKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Qy 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 7

US-10-103-327-8

; Sequence 8, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 424

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-103-327-8

Query Match 98.5%; Score 2288; DB 15; Length 424;
Best Local Similarity 100.0%; Pred. No. 8.2e-228;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QLRNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGWLHWRERFNCNLDCCQEPD 61
Db 1 QLRNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGWLHWRERFNCNLDCCQEPD 60
Qy 62 SCISEKLFMEAMELMVSEGKWDAGYEYLCIDDCWMAPOQDSEGRLOADPQRFPHGIRQLA 121
Db 61 SCISEKLFMEAMELMVSEGKWDAGYEYLCIDDCWMAPOQDSEGRLOADPQRFPHGIRQLA 120
Qy 122 NYVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 181
Db 121 NYVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 180
Qy 182 DGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 241
Db 181 DGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Qy 242 ILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 301
Db 241 ILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 300
Qy 302 HISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 361
Db 301 HISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 360

Qy 362 PRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 PRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 416

RESULT 8

US-10-103-327-4

; Sequence 4, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 428

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-103-327-4

Query Match 98.5%; Score 2288; DB 15; Length 428;
Best Local Similarity 100.0%; Pred. No. 8.3e-228;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QLRNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGWLHWRERFNCNLDCCQEPD 61
Db 1 QLRNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGWLHWRERFNCNLDCCQEPD 60
Qy 62 SCISEKLFMEAMELMVSEGKWDAGYEYLCIDDCWMAPOQDSEGRLOADPQRFPHGIRQLA 121
Db 61 SCISEKLFMEAMELMVSEGKWDAGYEYLCIDDCWMAPOQDSEGRLOADPQRFPHGIRQLA 120
Qy 122 NYVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 181
Db 121 NYVHSGKLGKGIYADVGNKTCAGFGSGFYDIDAQTFADGWGVDLLKFDGCGYCDSENL 180
Qy 182 DGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 241
Db 181 DGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Qy 242 ILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 301
Db 241 ILDWTSFNOERIVDVAGPGGWNDDMLVIGNFGLSWNOQVOTQMALWAIMAAPLFMSNDL 300
Qy 302 HISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 361
Db 301 HISPOKALLQDKVDIAINODPLGKGYQLRGDNFEVWERPLSGLAWAVAMINRQIEG 360
Qy 362 PRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 PRSYTTAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 416

RESULT 9

US-10-103-327-22

; Sequence 22, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-22

Query Match 96.2%; Score 2234.5; DB 15; Length 409;
Best Local Similarity 96.9%; Pred. No. 2.6e-222;
Matches 409; Conservative 0; Mismatches 0; Indels 13; Gaps 1;

QY 2 QLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCNDCEEPD 61
DB 1 QLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCNDCEEPD 60
QY 62 SCISEKLFMEMAELMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRQADPQRPFGHGIQOLA 121
DB 61 SCISEKLFMEMAELMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRQADPQRPFGHGIQOLA 120
QY 122 NYVHSGKLGKIYADVGNKTCAGFPGSGFYDIDAQTFADGWVDLLKFDGCGYCDSENLA 181
DB 121 NYVHSGKLGKIYADVGNKTCAGFPGSGFYDIDAQTFADGWVDLLKFDGCGYCDSENLA 180
QY 182 DGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYNHWRNFADIDDSWKSIS 241
DB 181 DGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYNHWRNFADIDDSWKSIS 240
QY 242 ILDTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNOQVOTOMALWAIMAAPLFMSNDLR 301
DB 241 ILDTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNOQVOTOMALWAIMAAPLFMSNDLR 300
QY 302 HISPOKALLQDKDVIAINQDPLGKQGYQLRGDNFVYWERPLSGLAWAVAMINRQEIIG 361
DB 301 HISPOKALLQDKDVIAINQDPLGKQGYQLRGDNFVYWERPLSGLAWAVAMINRQEIIG 360
QY 362 PRSYTTAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSLRINPTGTVLLQLSEKD 421
DB 361 PRSYTTAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLRSLRINPTGTVLLQLSEKD 420
QY 422 EL 423
DB 408 EL 409

RESULT 10
US-10-103-327-20
; Sequence 20, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 20
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-20

Query Match 95.3%; Score 2213; DB 15; Length 401;
Best Local Similarity 100.0%; Pred. No. 4.3e-220;

Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 RNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCNDCEEPDSC 63
DB 1 RNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCNDCEEPDSC 60
QY 64 ISEKLFMEMAELMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRQADPQRPFGHGIQOLANY 123
DB 61 ISEKLFMEMAELMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRQADPQRPFGHGIQOLANY 120
QY 124 VHSKGLKGIYADVGNKTCAGFPGSGFYDIDAQTFADGWVDLLKFDGCGYCDSENLAOG 183
DB 121 VHSKGLKGIYADVGNKTCAGFPGSGFYDIDAQTFADGWVDLLKFDGCGYCDSENLAOG 180
QY 184 YKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYNHWRNFADIDDSWKSIS 243
DB 181 YKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYNHWRNFADIDDSWKSIS 240
QY 244 DWTFSNOERIVDVAGPGGWNDDPMLVIGNFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHI 303
DB 241 DWTFSNOERIVDVAGPGGWNDDPMLVIGNFGLSWNOQVOTOMALWAIMAAPLFMSNDLRHI 300
QY 304 SPQAKALLQDKDVIAINQDPLGKQGYQLRGDNFVYWERPLSGLAWAVAMINRQEIIGPR 363
DB 301 SPQAKALLQDKDVIAINQDPLGKQGYQLRGDNFVYWERPLSGLAWAVAMINRQEIIGPR 360
QY 364 SYTTAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLR 404
DB 361 SYTTAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSLR 401

RESULT 11
US-10-165-060-4
; Sequence 4, Application US/10165060
; Publication No. US20030077806A1
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,060
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-165-060-4

Query Match 91.8%; Score 2131; DB 15; Length 398;
Best Local Similarity 100.0%; Pred. No. 1.3e-211;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 LONGLARTPTMGWLHWFRCMNCNDCEEPDSCISEKLFMEMAELMVSEGWKDGAGYEYLCI 91
DB 1 LONGLARTPTMGWLHWFRCMNCNDCEEPDSCISEKLFMEMAELMVSEGWKDGAGYEYLCI 60
QY 92 DDCWMAPOQDSEGRQADPQRPFGHGIQOLANYVHSGKLGKIYADVGNKTCAGFPGSGFY 151

Db 61 DDCWAPQDSEGRLOADPQRFPHGIRQLANYVHSKGLKGIYADVGNKTCAGFPSCFGY 120
 QY 152 YDIDAQTEADWGVLLKFDGCGYCDSENLEADGKHKMSLALNRTGRSIVTSCWPLYMWP 211
 Db 121 YDIDAQTEADWGVLLKFDGCGYCDSENLEADGKHKMSLALNRTGRSIVTSCWPLYMWP 180
 QY 212 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPDMVLIG 271
 Db 181 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPDMVLIG 240
 QY 272 NFGLSNQOVTQMALWAIMAAPLFMSNDLRHISPOAKALLODKDVIAINQDPLGKGYOL 331
 Db 241 NFGLSNQOVTQMALWAIMAAPLFMSNDLRHISPOAKALLODKDVIAINQDPLGKGYOL 300
 QY 332 ROGDNFEWERPLSGLAWAVAMINRQETGGPRSYTIAVASLGKGVACNPACFITQLLPVK 391
 Db 301 ROGDNFEWERPLSGLAWAVAMINRQETGGPRSYTIAVASLGKGVACNPACFITQLLPVK 360
 QY 392 RKLGFYEWTSRLRSHNPTGTVLLQL 417
 Db 361 RKLGFYEWTSRLRSHNPTGTVLLQL 386

RESULT 12

US-10-165-968-4
 ; Sequence 4, Application US/10165968
 ; Publication No. US200301138941
 ; GENERAL INFORMATION:
 ; APPLICANT: Seiden, Richard F
 ; APPLICANT: Borowski, Marianne
 ; APPLICANT: Kinoshita, Carol M
 ; APPLICANT: Treco, Douglas A
 ; APPLICANT: Williams, Melanie D
 ; APPLICANT: Schuetz, Thomas J
 ; APPLICANT: Daniel, Peter F
 ; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
 ; FILE REFERENCE: FABRY DISEASE (18082-001)
 ; CURRENT APPLICATION NUMBER: US/10/165,968
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: US/09/266,014
 ; PRIOR FILING DATE: 1999-03-11
 ; PRIOR APPLICATION NUMBER: 60/026,041
 ; PRIOR FILING DATE: 1996-09-13
 ; PRIOR APPLICATION NUMBER: 08/928,881
 ; PRIOR FILING DATE: 1997-09-12
 ; PRIOR APPLICATION NUMBER: PCT US97/16603
 ; PRIOR FILING DATE: 1997-09-12
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 398
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-165-968-4

Query Match 91.8%; Score 2131; DB 15; Length 398;
 Best Local Similarity 100.0%; Pred. No. 1.3e-211;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 32 LDNGLARTPTMGWLHWRFCMNCDCQEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 91
 Db 1 LDNGLARTPTMGWLHWRFCMNCDCQEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 60
 QY 92 DDCWAPQDSEGRLOADPQRFPHGIRQLANYVHSKGLKGIYADVGNKTCAGFPSCFGY 151
 Db 61 DDCWAPQDSEGRLOADPQRFPHGIRQLANYVHSKGLKGIYADVGNKTCAGFPSCFGY 120
 QY 152 YDIDAQTEADWGVLLKFDGCGYCDSENLEADGKHKMSLALNRTGRSIVTSCWPLYMWP 211
 Db 121 YDIDAQTEADWGVLLKFDGCGYCDSENLEADGKHKMSLALNRTGRSIVTSCWPLYMWP 180
 QY 212 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPDMVLIG 271

Db 181 QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPDMVLIG 240
 QY 272 NFGLSNQOVTQMALWAIMAAPLFMSNDLRHISPOAKALLODKDVIAINQDPLGKGYOL 331
 Db 241 NFGLSNQOVTQMALWAIMAAPLFMSNDLRHISPOAKALLODKDVIAINQDPLGKGYOL 300
 QY 332 ROGDNFEWERPLSGLAWAVAMINRQETGGPRSYTIAVASLGKGVACNPACFITQLLPVK 391
 Db 301 ROGDNFEWERPLSGLAWAVAMINRQETGGPRSYTIAVASLGKGVACNPACFITQLLPVK 360
 QY 392 RKLGFYEWTSRLRSHNPTGTVLLQL 417
 Db 361 RKLGFYEWTSRLRSHNPTGTVLLQL 386

RESULT 13

US-10-156-761-9014
 ; Sequence 9014, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 9014
 ; LENGTH: 682
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-9014

Query Match 29.4%; Score 683.5; DB 15; Length 682;
 Best Local Similarity 39.1%; Pred. No. 1.1e-61;
 Matches 168; Conservative 62; Mismatches 131; Indels 69; Gaps 17;
 QY 12 CALALRFLALVSWDIPGARLD---NGLARTPTMGWLHWRFCMNCDCQEPDSCISEK 67
 Db 32 CTAGLAAPALAA---PRAGVSPRPADSLALTPMGFNWNN---STGCRPE---FNED 80
 QY 68 LFMEAEMLVSEGWKDAGYEYLCIDDCWMAFQDSEGRLOADPQRFPHGIRQLANYVHSK 127
 Db 81 MVKGTADIFVEKGLKDAGYQVNLDDCWALPARDSNGKLVDPDPARFPGIKAVADYVHSK 140
 QY 128 GLKLGIIYADVGNKTC--AGPPGSGYDIDDAQTADWGVLLKFDGCGYCDSENLEADGYK 185
 Db 141 GLKLGIIYTSAGTKTCNEAGFPGALGHEYSDAQQFADWGVLYKYDNCNNOQGVDAKL-RYT 199
 QY 186 HMSLALNRTGRSIVYS-CEW---PLYMWPQKPNYTEIRQYCNHWRNFADIDDSWKSIS 240
 Db 200 TMRDALAKATRPVYSLCEWENKP---WEWA---SDVGQL---WRTGDISWSGML 249
 QY 241 SILDWTSFNQERIVDVAGPGWNDPDMVLIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
 Db 250 SILK---QNLPLAPYAGPGHNDPDMLEVNGSGMTDEYRTHFSWMSIMAAPLLIGSDL 305
 QY 301 RHISPOAKALLODKDVIAINQDPLGKGYOL-ROGDNFEWERPLSGLAWAVAMINR--- 356
 Db 306 RKASATFDILDNKRIVAVDDPDLGKQCTVLSSEGGRAWYTKEMKDG-SRAVALFNETDS 364
 QY 357 -----QETGPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHIN 408
 Db 365 AQRITTTAQAAGLPLKAHAYTLRDLWQHSRNTAGTISATVP-----AH-- 407

Qy 409 PTGTVLQLS 418
|||:::
Db 408 --GTVLVRVS 415

RESULT 14

US-10-156-761-8170
Sequence 8170, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8170
LENGTH: 536
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8170

Query Match 29.1%; Score 676.5; DB 15; Length 536;

Best Local Similarity 40.8%; Pred. No. 4.2e-61;
Matches 154; Conservative 52; Mismatches 132; Indels 39; Gaps 10;

Qy 10 LGCALALRFL-----ALVSWDIPGARALDNGLARTPTMGWLHWFEMCNLDCEEPDSC 63
|||:::
Db 11 LSAFAMLLLVLLAPALLVLRAPPQAQALDNGLARTPPMGWNNWNAFCGN----- 60
Qy 64 ISKLFMEMAELMVSEGWKADGVEYLCIDCWMAQRDSEGRLOADPQRPFGHRIOLANY 123
|||:::
Db 61 VTEOLVQOTADYLVSSGLKAGYQVNVNIDCWMTSARNSVQLVPDPVKPDPGIGSTAAY 120
Qy 124 VHSKGLGIYADVGNKTCAGFCGFCYDIDAQTFADGWVLLKFDGVCYCDLENLADG 183
|||:::
Db 121 VHSKGLGIYESAGTATCGYPGSIGHEQTDADSFASGWGVLYKYDNC---NHQNVPDQ 177
Qy 184 YKIMSL--ALNRTGRSTIVYS--CEWPL--YMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 239
|||:::
Db 178 QRYTAMRDALVNTGRPIVYSLCNWGLASVW-----TWGAGVCSNRTTDDINNVSTV 230
Qy 240 KSILDWTSFNOERIVDVAGGWNDDPMLVIGNPGLSNQOQVTOALMAAIPLFVNSD 299
|||:::
Db 231 VSIYK----ANVLAPYAKPGAWNDPMLVEYGN--GMSFTEDRSFSLWSEMAAPLIAGTD 285
Qy 300 LRHSPOKALLQDKVIALNODPLGKGYQLROGDNFEYWERPLSGLANAVAMINRQEI 359
|||:::
Db 286 LRKASATFLYGNKDVIAVDQDSLGKQGTVEVSSGGLHLVLTKPLANGDVSVLNFENS- 344
Qy 360 GGRPSYTIYAVASLKG 376
|||:::
Db 345 ---SAATITTSATAAGL 358

RESULT 15

US-10-156-761-9724
Sequence 9724, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9724
LENGTH: 658
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9724

Query Match 18.5%; Score 430; DB 15; Length 658;

Best Local Similarity 30.7%; Pred. No. 1.7e-35;
Matches 122; Conservative 52; Mismatches 159; Indels 64; Gaps 14;

Qy 8 LHLGCALALRFLALVSWDIPGARALDNGLARTPT--MGWLHWFEMCNLDCEEPDSCIS 65
|||:::
Db 26 LAUTAAATLLFVQ----PAPAQAATVTRQITVPSPAPGWSWNSFAAKID----- 71
Qy 66 EKLPMEMAEMLMVSEGWKADGVEYLCIDCWMAQRDSEGRLOADPQRPFGHRIOLANYVH 125
|||:::
Db 72 YSVIKQVDAFVAAGLPAAGVYTNIDEGWMOGTRDSAGNITVDTAEWPGMGSAITAYIH 131
Qy 126 SKGLKGIYADVGNKTC-----AGFP-----GSEGYDIDAQTFADGWVLLKFDGVCYCD 175
|||:::
Db 132 SKGLKGIYTDAGKDCGCGYYPTRGPAAPGSGSEHYDQDMLQFSTWGFDFVKVDCGGD 191
Qy 176 SLENL--ADGYKHMSLALNR----TGRSIVYS--CEWPLYMMPFQKPNYTEIROQCNHWRN 228
|||:::
Db 192 A-EGLDAAITYKSIDAVGRAAATTGRPLTISICNW-----GYQNPWNW 234
Qy 229 PADIDDSWKSITKSLDW-----TSFNOERIVDVAGGWNDDPMLVIGNPGLSNW 278
|||:::
Db 235 AAGQAPLWRTSTDIYYCNPQSPMTSLLSNFDTLHPTAQHTGYNDPDMLMVGMGDTAA 294
Qy 279 QOVTQMALWATMAAPLFMSNDLRHISPOAKALLQDKVIALNODPLGKGYQLROG-DNF 337
|||:::
Db 295 QNRTHMLWALSGLAPLAGNDLTMTSETAGILKNPEVIAVDQDSRGLQGVKVAEDTTGL 354
Qy 338 EVWERPLSGLA-WAVAMINRQEIIGGRPSYTIYAVASLG 373
|||:::
Db 355 QAYGKVLSGTCNRAVLLNR--TSAAHDIYVRWSDLG 389

Search completed: July 29, 2003, 16:29:35
Job time : 65.849 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:09:11 ; Search time 117.342 Seconds
(without alignments)
930.238 Million cell updates/sec

Title: US-09-993-059-18
Perfect score: 2322
Sequence: 1 MOLNPELHGCALALRFLA.....RSHNPTGTVLQLSEKDEL 423

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mhc:
- 8: sp_organelle:
- 9: sp_phase:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_rvirus:
- 16: sp_bacteriap:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1806	77.8	421	11	Q8BGZ6
2	1480.5	63.8	429	13	Q90WL7
3	1054	45.4	405	13	Q90744
4	1030	44.4	413	5	Q9VL27
5	1030	44.4	427	5	Q8MYV3
6	1007	43.4	417	5	Q9V7N9
7	990.5	42.7	415	11	Q88620
8	990.5	42.7	423	11	Q8R437
9	985.5	42.4	415	11	Q8VDK2
10	983.5	42.4	415	11	Q9QWR8
11	861	37.1	451	5	Q21801
12	729	31.4	413	10	Q9FWV8
13	716.5	30.9	434	10	Q9LYL2
14	716.5	30.9	437	10	Q8VXZ7
15	710	30.6	478	2	Q8VV86
16	695.5	30.0	380	10	Q9FUR8

17	695.5	30.0	409	10	Q9SP05
18	687.5	29.6	417	10	Q9FXT4
19	687.5	29.6	422	10	Q39811
20	684.5	29.5	396	10	Q8RX86
21	682	29.4	425	10	Q41100
22	680.5	29.3	410	10	Q9FT97
23	678.5	29.2	410	10	Q8RWB9
24	676	29.1	680	16	Q9S2C9
25	641	27.6	428	2	Q9Z4N5
26	633.5	27.3	438	3	Q9HF29
27	624.5	26.9	365	10	Q9FT98
28	622.5	26.8	438	3	Q9HG00
29	609	26.2	436	3	Q9UR20
30	598	25.8	396	3	Q93816
31	571.5	24.6	417	3	Q02402
32	557	24.0	469	3	Q99172
33	540.5	23.3	435	3	Q94221
34	525.5	22.6	470	3	Q9P4V4
35	524	22.6	443	3	Q9Y865
36	520	22.4	471	3	Q09187
37	516	22.2	444	3	Q92456
38	514	22.1	471	3	Q03647
39	510.5	22.0	454	3	Q96UV4
40	509	21.9	471	3	Q11129
41	505	21.7	439	3	Q13295
42	490	21.1	474	3	Q9UVD6
43	471	20.3	416	3	Q9HGJ7
44	444.5	19.1	416	3	Q9HGJ6
45	441.5	19.0	547	3	Q9HFB1

ALIGNMENTS

RESULT 1

Q8BGZ6	Q8BGZ6	PRELIMINARY;	PRT;	421	AA.
AC	Q8BGZ6;				
DT	01-MAR-2003 (Tremblrel. 23, Created)				
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)				
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)				
DE	Galactosidase.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;				
RX	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs."				
RL	Nature 420:563-573(2002).				
DR	EMBL; AK040081; BAC30508.1;				
DR	EMBL; AK054547; BAC35819.1;				
SO	SEQUENCE 421 AA; 7845 MW; 5F66772334014B6F CRC64;				
Query Match					
Best Local Similarity 77.8%; Score 1806; DB 11; Length 421;					
Matches 326; Conservative 41; Mismatches 50; Indels 0; Gaps 0;					
Qy	1	MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWHLHWRFCMNCDCOEEP	60		
Db	3	MKLLSRTRLVCELALCPALVFWSLGVRALDNGLARTPTMGWHLHWRFCMNCDCOEEP	62		
Qy	61	DSCLSEKLEFMEAEIMVSEGWKDAYEYLCITDDCWAPQDSEGRQLADPQFFHIGRL	120		
Db	63	DACISEQLFQMAELMVSQWGRDAGYDYLCITDDCWAPERDSKGRQLADPQFFSGIKHL	122		
Qy	121	ANYVHSGKLKLGIVADYGNKTCAGFPGSFGYDIDAQTFADWGVDLKFDGCYCDSELENL	180		

```
Db 123 ANTVHSGKGLGIYADVGNKTKAGFGSGSGYDIDAQTFADWGVDLKFDGCHCDSVSL 182
QY 181 ADGKHMALNRTGRSIVYSCWPLYMPPQKPNYTEIROYCNHNRNFADIDDSKSIK 240
Db 183 ENGIYKYMALNRTGRSIVYSCWPLYLRFPHPNYTDIOYCNHNRNEDDVYDSHESIK 242
QY 241 SILDWTSFNOERIVDVAGPGWNDPMLYIGNFGLSNQOQVTOALWAIMAAPLFMSNDL 300
Db 243 NILSWFWVYQKEIVEVAGPGWNDPMLYIGNFGLSWDQOQVTOALWAIMAAPLLMSNDL 302
QY 301 RHLSPOAKALLQDKVIAINODPLGKGYLROGDNFVWERPLSLGLAWAVAMINRQEI 360
Db 303 RQISSQAKALLQNDKVIATNODPLGKGYCFRKNHIEVWERPLSLGLAWAVARNQEI 362
QY 361 GPRSYTIASVSLGKGVACVACITQLLPVKRLGFYEWTSRLRSHINPTGTIVLLQL 417
Db 363 GPCPYTIQISSLGRGLACNPGCIITQLLPEKVLHGFYEWTLTKTRVNSGTVLFL 419

RESULT 2
Q90WL7 PRELIMINARY; PRT; 429 AA.
AC Q90WL7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Alpha galactosidase a.
GN ACAL.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OC NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Goode D., Elgar G.;
RT "Three way comparative genomic analysis of the BTK locus between man,
mouse and Fugu."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ290422; CAC44626.1;
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 429 AA; 48480 MW; 536A826D520CF01C CRC64;

Query Match 63.8%; Score 1480.5; DB 13; Length 429;
Best Local Similarity 68.4%; Pred. No. 1.4e-125;
Matches 271; Conservative 50; Mismatches 72; Indels 3; Gaps 2;

QY 20 ALVSWDIPGARDLNGRLARTPTMGLHWRFCNLDQCEPDSCISEKLFMEAEMLVSE 79
Db 7 ALVFLGPAEALDNLGALTPTMGLHWRFCNLDQCEPDSCISEKLFMEAEMLVSE 66
QY 80 GWKDAGYEYLCIDCWAPQRDSGRQLQADPQRFPHGIRQLANYVHSGKGLGIYADVGN 139
Db 67 GWKEAGYEYVCIDCWAPQRDSGRQLQADPQRFPHGIRQLANYVHSGKGLGIYADVGN 126
QY 140 KTCAGPGSGYGYDDAQTFAWGVDLKFDGCVCDSENLDAGYKHMSLALNRTGRSIV 199
Db 127 KTCAGPGSGYGYDDAQTFAWGVDLKFDGCVCDSENLDAGYKHMSLALNRTGRSIV 185
QY 200 YSCWPLYMPPQKPNYTEIROYCNHNRNFADIDDSKSIKSLDWTSENOERIVDVAGP 259
Db 186 YSCWPLYMPPQKPNYTEIROYCNHNRNFADIDDSKSIKSLDWTSENOERIVDVAGP 245
QY 260 GGNNDPMLYIGNFGLSNQOQVTOALWAIMAAPLFMSNDLRHSPOAKALLQDKVIAI 319
Db 246 GGNNDPMLYIGNFGLSHDQESOMALWAIMAAPLFMSNDLRHSPOAKALLQDKVIAI 305
```

```
QY 320 NODPLGKGYLROGDNFVWERPLSLGLAWAVAMINRQEIIGRPSRTYIAVASLGKGVACN 379
Db 306 NODPLGKGYLROGDNFVWERPLSLGLAWAVAMINRQEIIGRPSRTYIAVASLGKGVACN 363
QY 380 PACFITQLLPVKRLGFYEWTSRLRSHINPTGTIVLL 415
Db 364 PACNVTQILPHYKDMGVQPMFSGMVVRVNSPGTVLL 399

RESULT 3
Q90744 PRELIMINARY; PRT; 405 AA.
AC Q90744;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Alpha-N-acetylgalactosaminidase (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94060104; PubMed=8241271;
RX Davis M.O., Hata J., Smith D., Walker J.C.;
RT "Cloning and sequence of a chicken alpha-N-acetylgalactosaminidase
gene."
RL Biochim. Biophys. Acta 1216:296-298(1993).
DR EMBL; L18754; AAA16614.1;
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
FT NON_TER 1
SQ SEQUENCE 405 AA; 45615 MW; E1EC0061739C305C CRC64;

Query Match 45.4%; Score 1054; DB 13; Length 405;
Best Local Similarity 52.8%; Pred. No. 5.8e-87;
Matches 207; Conservative 58; Mismatches 119; Indels 8; Gaps 6;

QY 32 LONGLARTPTMGLHWRFCNLDQCEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 91
Db 1 LENGLARTPTMGLHWRFCNLDQCEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 60
QY 92 DCWMAQRDSGRQLQADPQRFPHGIRQLANYVHSGKGLGIYADVGNKTCAGFGP-SFG 150
Db 61 DCWMAQRDSGRQLQADPQRFPHGIRQLANYVHSGKGLGIYADVGNKTCAGFGP-SFG 120
QY 151 YXDIDAQTFAWGVDLKFDGCVCDSENLDAGYKHMSLALNRTGRSIVYSCWPLYMWP 210
Db 121 RVEDAQTFAWGVDLKFDGCVCDSENLDAGYKHMSLALNRTGRSIVYSCWPLYMWP 179
QY 211 F-QKNPYTIROYCNHNRNFADIDDSKSIKSLDWTSENOERIVDVAGPGWNDPMLV 269
Db 180 LPPKVNVTLLGELCNLRNVDIQQSDWSVLSVDWFFTNQDVLQFPAGPGHNDPMLI 239
QY 270 IGNFGLSNQOQVTOALWAIMAAPLFMSNDLRHSPOAKALLQDKVIAINODPLGKGY 329
Db 240 IGNFGLSNQOQVTOALWAIMAAPLFMSNDLRHSPOAKALLQDKVIAINODPLGKGY 299
QY 330 Q-LROGDNFVWERPLSLGLAWAVAMINRQEIIGRPSRTYIAVASLGKGVACNPAFITOLL 388
Db 300 RRIKESGHEVLERPLSLGLAWAVAMINRQEIIGRPSRTYIAVASLGKGVACNPAFITOLL 355
QY 389 PVKRLGFYEWTSRLRSHINPTGTIVLLQLSEK 420
Db 356 YSKKIISGLKTDGNTFTVINPSGVVMWYLCPK 387
```

```

RESULT 4
Q9VL27 ID Q9VL27 PRELIMINARY; PRT; 413 AA.
AC Q9VL27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C5731 protein.
GN C5731.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Venter L.M., Zhang Q., Chen L.X.,
RA Ruan K.H., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C.C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier A.S., Fleischmann W.,
RA Fosler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragass V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RP Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```

```

[3]
RN RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G.F., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003627; AAF52871.2;
DR FlyBase; FBgn0032192; CG5731.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melbiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 413 AA; 47047 MW; 9ADE417E592782D CRC64;
Query Match 44.4%; Score 1030; DB 5; Length 413;
Best Local Similarity 46.7%; Pred. No. 8.9e-85;
Matches 191; Conservative 81; Mismatches 125; Indels 12; Gaps 5;
Qy 14 LALREFALVSDWIPGARALDNLGARTPTMGWLHFERMCNLDQCEPDSCISEKLFEMMA 73
Db 2 LATWILLISSTF--GKCLDNLGARTPTMGWLHFERMCNLDQCEPDSCISEKLFEMMA 73
Qy 74 ELMYSEGKADGAYEYLCLIDCCWAPQDSRGLAQDPQRFPHGIRQLANYVHSKGLGI 133
Db 60 DIVVADGVASVGYEYINIDCCWLEKHSRSHDCKLVADRKRFPNGIKALSDYHSRGLKFGI 119
Qy 134 YADVGNKTCAGFCGSGFYDDAQTQFADMGVLLKFDCCYCDLSLENLADGKXHSNLNR 193
Db 120 YEDYGNKTCAGYGIIGYEKDAQFADMGVLLKFDCCYCDLSLENLADGKXHSNLNR 178
Qy 194 TGRSIVSCEWPLY-MMPFQKNYETROYCNHNRNFADIDDSNKSISILDTWTSFNOER 252
Db 179 TGKSNVSCSWPVYQYVAGIQPNYSATQHCNLRNYDDIDDSNKSISILDTWTSFNOER 238
Qy 253 IVDVAGPGWMDPMLVIGNFGSLWNOQVOTQMALWAIMAPLFMSNDRHISPOKALLQ 312
Db 239 IAPNAGPGWMDPMLVIGNFGSLWNOQVOTQMALWAIMAPLFMSNDRHISPOKALLQ 298
Qy 313 DKQVIAINQDPLGKQGLRGDNFVWVERPESGL-----AWAVAMINRQEIIGPRSYTI 367
Db 299 NRKTIIVDQDPLGKQGLRGDNFVWVERPESGL-----AWAVAMINRQEIIGPRSYTI 358
Qy 368 AVASLGKGVACNPACFITQLLPVKKRKLGFYEWTSRLSRHINPTGTVLLQ 416
Db 359 TLKELG---LINFSGYVEDLYENVYGVLPNTKIRKVKVNPSPGVVMLK 404
RESULT 5
Q8MY3 ID Q8MY3 PRELIMINARY; PRT; 427 AA.
AC Q8MY3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RE47112p.
GN CG5731.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Capleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarini H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.,
 RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AY113489; AAM29494.1;
 DR FlyBase; FBgn0032192; CG5731.
 DR InterPro; IPR002241; Glyco_hydro_27.
 DR InterPro; IPR000111; Glyco_hydro_GHD.
 DR PRINTS; PR00740; GLHYDRLA27.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 SQ SEQUENCE 427 AA; 48678 MW; 72EDBA7A22346FFB CRC64;
 Query Match 44.4%; Score 1030; DB 5; Length 427;
 Best Local Similarity 46.7%; Pred.No 9.4e-85;
 Matches 191; Conservative 81; Mismatches 125; Indels 12; Gaps 5;
 QY 14 LAURFALVSWDIPGARALDNLGARTPTMGWLHWRFCMLDQEPDSCISEKLEWMA 73
 Db 2 LATLWILLSSSTP--GKCLDNLGLAKTPPMGLSWERFCNCTCVNDPDCNISQLFQTMT 59
 QY 74 ELMVSGKWDAGVEYLCIDCCWMAPODSEGRLOADPQRFPHGIROLANYVHSGKLGI 133
 Db 60 DIVVADGYASVGEYINIDCWLEKHSRHDGKLVADKRFPNGIKALSYIHSRGLKFGI 119
 QY 134 YADVGNKTCAGFGSGFYDIDAQTEADGCVDLLKFDGVCDSLENLADGKHMSLALNR 193
 Db 120 YEDYGNVTCAGYPIIGIYEKQALQADNVDYKLDGCVALPYD-MDGYSTFGRLNLS 178
 QY 194 TGRSIYVSEWPLY-WMPFQKPNYETIQYCNHWRNFADIDDSWKSIKSILDWTSNQER 252
 Db 179 TCKSMYVCSWPYQIYAGIYQNYSAIQTHCNLWRNYDDTQDSWASVENIIDYGNQDV 238
 QY 253 IVDVAGPGGWNDDMLVIGNFLGSNQOVTOMALWAIMAAPLMSNDLHISPOAKALIQ 312
 Db 239 IAPNAGPGHWNDDMLIIGNFLGSYEQAKTQFAIWSILAAPLMSVDLRTIRPQFKHILQ 298
 QY 313 KXDVTAINQDPLGKQYQLRQGNDFEWRPLSGL-----AWAVAMINRQEIGGPRSYTI 367
 Db 299 NRKIIADVDDPLGIQGRRIYKHKIEWSRPGIPLQYFYSAIAFVNRRTDGTSPDSIVS 358
 QY 368 AVASLGKGVACNPACITQLLPVKRKLGYEWSRLRSRSHINPTGVLLQ 416
 Db 359 TLKELG--LINFSGYRVEDLVNDYGVLYPNTKIKVKVNPFGVVMVK 404
 RESULT 6
 QYV7N9 PRELIMINARY; PRT; 417 AA.
 AC QYV7N9; Q870B5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG7997 protein (Ldl3649p).
 GN CG7997.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Rosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G.S., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;


```

DE Alpha-N-acetyl-galactosaminidase (Fragment).
GN NAGA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RX MEDLINE=21363810; PubMed=11471062;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-throughput sequence identification of gene coding variants
RT within alcohol-related QTLs.";
RL Mamm. Genome 12:657-663(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY079439; AAL87527.1;
DR MGI: 1261422; Naga.
DR InterPro: IPR000111; Glyco_hydro_GHD.
DR ProDom: PD002572; Glyco_hydro_GHD; 1.
DR PROSITE: PS00512; ALPHA_GALACTOSIDASE; 1.
FT NON_TER 1
FT SEQUENCE 423 AA; 48130 MW; D8A1AB1248C8D09B CRC64;

Query Match 42.7%; Score 990.5; DB 11; Length 423;
Best Local Similarity 55.8%; Pred. No. 3.5e-81;
Matches 191; Conservative 51; Mismatches 91; Indels 9; Gaps 6;

QY 19 LALVSWDIPGARALDNLGARTPTMGWLHWRPNCNLDCEPDSCISEKLFMEMAELMVS 78
DB 17 LALVAQVL----MLENGLLRTPPMGWLAWERFCNIDCEDPKNCISELFLMEMADRLAQ 72
QY 79 EGWKDAGYEYLCIDCCWMAQRDSEGRLOADPQRPFGHIGRLQANYVHSGKLGKIYADV 138
DB 73 DGWRDLGYVYLNDICWIG-GRDASGRLLPDKRPFGHIAFLADYAHSLGLGIYEDMG 131
QY 139 NKTACGFPG-SFGYYDIDAQTFADMGVDLLKFDGVCYDLSLENLADGKIMSLALNRTGRS 197
DB 132 KMTCMGYPGTTLDKVELDAETAEWKVDMLKLDGCFSSSRER-AEGYPKMAAALNATGRP 190
QY 198 IYVSCWPLYMPPF-QKPNYTEIRQYCNHWRNFADIDDSWKSITKSLDWTSTNQRIVDV 256
DB 191 IAFSCSWPAYEGGLPPKVNYTEVSRVNCNLWRNYKDIQDSWKSVLSTLDVFRHQDVLPV 250
QY 257 AGPGWNDDPMLVIGNFGLSNQOQVTQMALWAIMAPLWMSNDLRHSIPQAKALLQDKDV 316
DB 251 AGPGHWNDDPMLLIGNFGLSFDQRAQMALWTVLAAPLLMSTDLRTISPNQNDILQNLPLM 310
QY 317 IAINQDPLGKQGYQ-LRQGNDFEVRPLSLGLAWAVAMINRQ 357
DB 311 IKINQDPLGIQGRRLILKSKSHIEVFKRYLSNQASALVFFSRR 352

RESULT 9
Q8VDK2 PRELIMINARY; PRT; 415 AA.
AC Q8VDK2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE N-acetyl galactosaminidase, alpha.
GN NAGA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

Alpha-N-acetyl-galactosaminidase (Fragment).
GN NAGA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC021631; AAH21631.1;
DR MGI: 1261422; Naga.
DR InterPro: IPR000241; Glyco_hydro_27.
DR InterPro: IPR000111; Glyco_hydro_GHD.
DR Pfam: PF02065; Melibiase; 1.
DR PRINTS: PR00740; GLHYDRLASE27.
DR ProDom: PD002572; Glyco_hydro_GHD; 1.
DR PROSITE: PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 415 AA; 47218 MW; 82FA92A962060588 CRC64;

Query Match 42.4%; Score 985.5; DB 11; Length 415;
Best Local Similarity 55.8%; Pred. No. 9.7e-81;
Matches 191; Conservative 50; Mismatches 92; Indels 9; Gaps 6;

QY 19 LALVSWDIPGARALDNLGARTPTMGWLHWRPNCNLDCEPDSCISEKLFMEMAELMVS 78
DB 9 LALVAQVL----MLENGLLRTPPMGWLAWERFCNIDCEDPKNCISELFLMEMADRLAQ 64
QY 79 EGWKDAGYEYLCIDCCWMAQRDSEGRLOADPQRPFGHIGRLQANYVHSGKLGKIYADV 138
DB 65 DGWRDLGYVYLNDICWIG-GRDASGRLLPDKRPFGHIAFLADYAHSLGLGIYEDMG 123
QY 139 NKTACGFPG-SFGYYDIDAQTFADMGVDLLKFDGVCYDLSLENLADGKIMSLALNRTGRS 197
DB 124 KMTCMGYPGTTLDKVELDAETAEWKVDMLKLDGCFSSSRER-AEGYPKMAAALNATGRP 182
QY 198 IYVSCWPLYMPPF-QKPNYTEIRQYCNHWRNFADIDDSWKSITKSLDWTSTNQRIVDV 256
DB 183 IAFSCSWPAYEGGLPPKVNYTEVSRVNCNLWRNYKDIQDSWKSVLSTLDVFRHQDVLPV 242
QY 257 AGPGWNDDPMLVIGNFGLSNQOQVTQMALWAIMAPLWMSNDLRHSIPQAKALLQDKDV 316
DB 243 AGPGHWNDDPMLLIGNFGLSFDQRAQMALWTVLAAPLLMSTDLRTISPNQNDILQNLPLM 302
QY 317 IAINQDPLGKQGYQ-LRQGNDFEVRPLSLGLAWAVAMINRQ 357
DB 303 IKINQDPLGIQGRRLILKSKSHIEVFKRYLSNQASALVFFSRR 344

RESULT 10
Q9QWR8 PRELIMINARY; PRT; 415 AA.
AC Q9QWR8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha-N-acetyl galactosaminidase (EC 3.2.1.49).
GN NAGA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129 SV; TISSUE=Liver;
RX MEDLINE=98267200; PubMed=9602128;
RA Hermann T., Schindler D., Tabé H., Onodera O., Igarashi S.,
RA Polack A., Zehnfenning D., Tsuji S.;
RT "Molecular cloning, structural organization, sequence, chromosomal
RT assignment, and expression of the mouse alpha-N-
RT acetyl galactosaminidase gene.";
RL Gene 211:205-214 (1998).
DR EMBL: AJ223966; CAAL1703.1;
DR MGI: 1261422; Naga.
DR InterPro: IPR00241; Glyco_hydro_27.
DR Pfam: PF02065; Melibiase; 1.
DR PRINTS: PR00740; GLHYDRLASE27.
DR ProDom: PD002572; Glyco_hydro_GHD; 1.
DR PROSITE: PS00512; ALPHA_GALACTOSIDASE; 1.
DR Glycosidase; Hydrolase.

```

```
SQ SEQUENCE 415 AA; 47121 MW; A2CA292BA91AB347 CRC64;
Query Match 42.4%; Score 983.5; DB 11; Length 415;
Best Local Similarity 55.6%; Pred. No. 1.5e-80;
Matches 190; Conservative 51; Mismatches 92; Indels 9; Gaps 6;

QY 19 LALVSWDIPGARALDNGIARTPTMGWLHWFRCMNLDCQEPDSCISEKLFEMEMALMVS 78
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 9 LALVAQVL-----MLENGLLRTPPMGLAWERFRCNIDCEDPKNCISDGLFEMEMADRLAQ 64
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
QY 79 EGKWDAGVEYLICIDCWAPORSEGRLOADPQRPFGHGIROLANYVHSKGLKGIYADVG 138
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 65 DGRDLGVYINIDCWIG-GRDASGRILPDKRPHGIAFLADYAHSLGKLGITYEDMG 123
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
QY 139 NKTCAFPFG-SFGYDIDAQTFADMGVLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRS 197
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 124 KWTGCGPYGTTLDKVELDAETFAEWKVDMLKLDGCFSSRRER-AEGYPKMAAALNATGRP 182
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
QY 198 IYVSCWPLYMWPF-QRPNTYETIRQYCNHWNFRFADIDDSWKSISILDWTSFNQERIVDV 256
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 183 IAFSCSWPAYEGGLPPKYNVTEVSRVNCNLMWNYKDIDQSWKSIVLSILDWFRHQDVLPQV 242
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
QY 257 AGPGGWNDDMLVIGNFGLSNNOQVOTOMALWAIMAAPLMSNDRHISPOAKALLODKDV 316
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 243 AGPGHWNDDMLLIGNFGLSDESRAQNALMTVLAAPLMTSDTURTISPPQNNDILONPLM 302
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
QY 317 IAINQDPLGKQGYO-LROGDNFVYWRPLSLGLAWAVAMINRQ 357
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 303 IKINQDPLGIGORRLTKSKSHIEVFKRYLSNQASALVFESRR 344
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :

RESULT 11
Q21801 PRELIMINARY; PRT; 451 AA.
ID Q21801
AC Q21801
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE R07B7.11 protein.
GN R07B7.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z75955; CAB00120.1; -.
DR WormPep; R07B7.11; CE06273.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 451 AA; 51424 MW; F7BB2B0D5D5C6419 CRC64;

Query Match 37.1%; Score 861; DB 5; Length 451;
Best Local Similarity 43.8%; Pred. No. 2.1e-69;
Matches 174; Conservative 61; Mismatches 148; Indels 14; Gaps 6;

QY 28 GARALDNGIARTPTMGWLHWFRCMNLDCQEPDSCISEKLFEMEMALMVSSEGKWDAGYE 87
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
Db 12 GAFCLDNLGLRTPPMGWSWTAFCYCEIDCVKHTGTCINEQLYKDMADOLVSGGYDKVGYK 71
   ||||| : ||||| ||||| ||||| ||||| ||||| ||||| ||||| :

QY 88 YLCIDDCWMAFORSEGRLOADPQRPFGHGIROLANYVHSKGLKGIYADVGNKTCAGPPG 147
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 72 SVHIDDCWSEMERDGHGILVANKTRFPGSKMALAKYHMDRGLKFGIYEDYGTCTCGGYPG 131
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 148 SFGYDIDAQTFADMGVLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCWPLY 207
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 132 SYKHEKVDQAQFAAWDWDYLLKLDGCGNIDQ-AMMPGIGYPLFEKELNETGRPIMYSWPA 190
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 208 MWPFQKP---NYTEIROYCNHWNFRFADIDDSWKSISILDWTSFNQERIVDVAGPGWND 264
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 191 L-IDHPELVNLYGKYCNTRNFRDINSWKSISIIISYYDKMDQKHIPHTHGFQKWH 248
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 265 PDMVLVIGNFGLSWNOQVOTOMALWAIMAAPLMSNDRHISPOAKALLODKDVIAINQDPL 324
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 249 PDMVLVIGNKGTLDMSISQFTVWCISAPLMSNDRITIGDSFKDLVKNKEAIKINQDPL 308
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 325 GKQGYOLROGDNFVYWRPLSLGLAWAVAMINRQEIIGGPRSYTIAVASLGGKGVACN 379
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 309 GIMGLINKNSDIDGIVYKQITPFGKDKKSFAYLNRNEKEGYKRIETQLASIG---LTD 365
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 380 PACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQ 416
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 366 PAGYVVDIWSHVDLGLLRSGDSIVVSIAPAGSVFFR 402
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :

RESULT 12
Q9FWV8 PRELIMINARY; PRT; 413 AA.
ID Q9FWV8
AC Q9FWV8
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative alpha-galactosidase.
GN OSJNBA0051D19.18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Hsiao C., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBA0051D19 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC023240; AAG13536.1; -.
DR Gramene; Q9FWV8; -.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
SQ SEQUENCE 413 AA; 44665 MW; 03FB38A25F8FBCE2 CRC64;

Query Match 31.4%; Score 729; DB 10; Length 413;
Best Local Similarity 46.8%; Pred. No. 1.7e-57;
Matches 155; Conservative 43; Mismatches 103; Indels 30; Gaps 8;

QY 32 LDNGIARTPTMGWLHWFRCMNLDCQEPDSCISEKLFEMEMALMVSSEGKWDAGYEYLCI 91
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 39 LANGIARTAPPMGWSNHNHFACDGN-----GEDVIRETADALVSTGLAAAGKYKYN 89
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 92 DDCWMAFORSEGRLOADPQRPFGHGIROLANYVHSKGLKGIYADVGNKTCAGF-PGSFG 150
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
Db 90 DDCWMAFORSEGRLOADPQRPFGHGIROLANYVHSKGLKGIYADVGNKTCAGF-PGSFG 149
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
QY 151 YYDIDAQTFADMGVLLKFDGCGYCDLSLENLADGKYHMSLALNRTGRSIVYSCW---P 205
   : ||||| : ||||| ||||| ||||| ||||| ||||| ||||| :
```


Db 209 MRDALNATGRSIFYSLCEWGVDDPALW-----AKEVGNWRTTDDINDTWASMTTI 259
QY 243 LD----WTSFQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTCMALWMAAPLFMSN 298
Db 260 ADLNKNAAY-----AGPGGNDPDLMEIGNGMTYEEYRGHFSIWMKAPLLIGC 311
QY 299 DLRHISPOAKALLODKVDIAINQDPLGKQGYQLR---QGDNFVWERPGLSLAWAVAMIN 355
Db 312 DVNRNMTAETLILSKKEIIAVNQDPLGVQGRKIQANGENDCQVWVGSLGDRMVVALWN 371
QY 356 R 356
Db 372 R 372

RESULT 15

Q8VV86 PRELIMINARY; PRT; 478 AA.
AC Q8VV86;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Alpha-galactosidase.
GN AGAA.
OS Clostridium josui.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID:1499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FERM P-9684;
RX MEDLINE=21623568; PubMed=11751843;
RA Jindou S., Karita S., Fujino T., Hayashi H., Kimura T.,
RA Sakka K., Ohmura K.;
RT "alpha-Galactosidase Aga27A, an Enzymatic Component of the Clostridium
J. Bacteriol. 184:600-604(2002)."
RL EMBL; AB025362; BAB83765.1; -;
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002241; Glyco_hydro_27.
DR pfam; PF00404; Dockerin_1; 2.
DR pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 1.
DR PROSITE; PS00018; EF_HAND; 2.
SQ SEQUENCE 478 AA; 52163 MW; DDBF19ACDDBFA2C CRC64;

Query Match 30.6%; Score 710; DB 2; Length 478;
Best Local Similarity 43.8%; Pred. No. 1.le-55;
Matches 162; Conservative 48; Mismatches 118; Indels 42; Gaps 12;
QY 13 ALALRFLALVSWDIPGA-----RALDNGLARTPTMGWLHWRFCMNLDCOEEPDSCIS 65
Db 9 ALAILTWVFIS-PPASVLNFNQVWALNGLGLTPPGMGNWNIF-----GGDIN 57
QY 66 EKLFMEMAEMLVSEGKDGAGYEYLCIDCWNA-PORSEGRLOADPQRPFGIROLANYV 124
Db 58 EDKIEIADAMVTTGMKDAGYEYVNLDDNMANPARDANGKLIPDKRFPSPGMKALADYI 117
QY 125 HSKGLKGLIYADVGNKTCAGFP--GSFGYDIDAGTFADWGVDDLKFPDCCYCDSENLA 182
Db 118 HSKGLKFGIYDGRVTTCCNIPQSGSQSYEEQDAKTFAEWGLDYLKYDNCASDS--NLQA 175
QY 183 GYKHSLSALNRTGRSIVSCWPLYMPPFQPNYTEIRQYC-NHWRNFADIDDSWKSIS 241
Db 176 GYEKMRDALLTKGRIFYS-----ICCWTFAGPWWVD-----CNSWRTTGDIDSWSGIIR 227
QY 242 ILDTWTSFQERIVDVAGPGGNDPDLVIGNFGLSWNOQVTCMALWMAAPLFMSNDLR 301

Db 228 NIDENS-----KSAAYAGPGHWNDDPMLLEVGNNGNMTDTEYKAHFSMCMMAAPLIAAGNDLR 283
QY 302 HISPOAKALLODKVDIAINQDPLGKQGYQLROGDNFEYWERPL--SGLAWAVAMTNROEI 359
Db 284 NWTATKKEILLTNKEVIAIDQDAAGVQGTKVSSSGELEVWCKPLGTDGTTKAVALLNR--- 340
QY 360 GGPRSYTIAY 369
Db 341 -GATSADITV 349

Search completed: July 29, 2003, 16:17:11
Job time : 120.342 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 15:40:04 ; Search time 6.49664 Seconds
(without alignments)
586.370 Million cell updates/sec

Title: US-09-993-059-30

Perfect score: 118

Sequence: 1 TSRLRSHINPTGVLLQLSEKDEL 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	100.0	24	23 AAE19259	Human recombinant
2	118	100.0	423	23 AAE28213	Human rGAL-12R pro
3	106	89.8	28	23 AAE19257	Human recombinant
4	106	89.8	427	23 AAE28211	Human rGAL-8R prot
5	104	88.1	32	23 AAE19255	Human recombinant
6	104	88.1	431	23 AAE28209	Human rGAL-4R prot
7	102	86.4	36	23 AAE19253	Human recombinant
8	102	86.4	435	23 AAE28207	Human WT rGAL-AR p
9	89	75.4	18	23 AAE19258	Human recombinant

10	89	75.4	22	23 AAE19256	Human recombinant
11	89	75.4	26	23 AAE19254	Human recombinant
12	89	75.4	30	23 AAE19252	Human recombinant
13	89	75.4	92	15 AAR53771	Protein A domain E
14	89	75.4	386	22 AAB47203	Mutant delta 12 al
15	89	75.4	387	22 AAB47202	Mutant delta 11 al
16	89	75.4	388	22 AAB47201	Mutant delta 10 al
17	89	75.4	389	22 AAB47200	Mutant delta 9 al
18	89	75.4	390	22 AAB47199	Mutant delta 8 al
19	89	75.4	391	22 AAB47198	Mutant delta 7 al
20	89	75.4	392	22 AAB47197	Mutant delta 6 al
21	89	75.4	393	22 AAB47196	Mutant delta 5 al
22	89	75.4	394	22 AAB47195	Mutant delta 4 al
23	89	75.4	396	22 AAB47194	Mutant delta 2 Alp
24	89	75.4	398	19 AAW37988	Amino acid sequenc
25	89	75.4	398	21 AAB15387	Human alpha-Gal A
26	89	75.4	398	21 AAB15845	Amino acid sequenc
27	89	75.4	398	22 AAB47193	Alpha-galactosidas
28	89	75.4	398	23 AAE24540	Human alpha-gal A
29	89	75.4	417	23 AAE28212	Human rGAL-17
30	89	75.4	421	23 AAE28210	Human rGAL-8 prote
31	89	75.4	424	23 AAE28208	Human rGAL-4 prote
32	89	75.4	428	14 AAR31898	Human alpha-galact
33	89	75.4	429	11 AAR07305	Precursor form of
34	89	75.4	429	15 AAR63234	Human alpha-galact
35	89	75.4	429	15 AAR53765	Alpha-galactosidas
36	89	75.4	429	16 AAR70207	Alpha-galactosidas
37	89	75.4	429	18 AAW40801	Alpha-galactosidas
38	89	75.4	429	23 AAE28206	Human WT rGAL-A pr
39	89	75.4	430	24 ABG74455	Human alpha-galact
40	89	75.4	478	20 AAY48575	Human breast tumou
41	68	57.6	381	22 AAB47204	Mutant delta 17 al
42	54	45.8	203	21 AAB15981	E. coli proliferat
43	52	44.1	2515	22 ABB64427	Drosophila melanog
44	51	43.2	313	22 ABB61952	Drosophila melanog
45	51	43.2	316	22 ABB67340	Drosophila melanog

ALIGNMENTS

RESULT 1
AAE19259 standard; peptide; 24 AA.

XX AAE19259;

XX 21-MAY-2002 (first entry)

XX Human recombinant alpha-galactosidase A modified peptide, rGal-12R.

XX Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;

XX enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;

XX Fabry disease; lysosomal storage disease; human.

XX Homo sapiens.

XX Synthetic.

XX WO200208404-A2.

XX 31-JAN-2002.

XX 26-JUL-2001; 2001WO-US24111.

XX 26-JUL-2000; 2000US-0626127.

XX (LARG-) LARGE SCALE BIOLOGY CORP.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2002-195873/25.

XX New glucocerebrosidase and alpha-galactosidase having a

PT post-translational modification, useful in enzyme replacement therapy
 PT for treating lysosomal storage diseases, e.g. Gaucher disease,
 PT Niemann-Pick disease, Fabry disease
 XX
 PS Example 11; Fig 5; 102pp; English.
 XX
 CC The invention relates to production of lysosomal enzymes in plants
 CC by transient expression. The invention particularly relates to
 CC glucocerebrosidase and alpha-galactosidase having a post-translational
 CC modification. The enzymes are useful in enzyme replacement therapy for
 CC treating lysosomal storage diseases, (e.g. Gaucher disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease), in researches for
 CC developing new approaches to medical treatment of lysosomal storage
 CC diseases and in industrial processes involving enzymatic substrate
 CC hydrolysis. The present sequence is human recombinant alpha-
 CC galactosidase-A C-terminal modified peptide.
 XX
 SQ Sequence 24 AA;
 Query Match 100.0%; Score 118; DB 23; Length 24;
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TSLRSHINPTGVLLQSEKDEL 24
 Db 1 TSLRSHINPTGVLLQSEKDEL 24
 RESULT 2
 AAE28213
 ID AAE28213 standard; Protein; 423 AA.
 XX
 AC AAE28213;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 DE Human rGAL-12R protein.
 XX
 KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
 KW therapeutic; rGAL-12R.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Cleavage-site 401..402
 FT /note= "CTPP cleavage site"
 XX
 PN US2002088024-A1.
 XX
 PD 04-JUL-2002.
 XX
 XX 13-NOV-2001; 2001US-0993059.
 PF
 XX 26-JUL-2000; 2000US-0626127.
 PR
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 XX
 XX WPI; 2002-691656/73.
 DR N-PSDB; AAD45226.
 XX
 XX Novel human alpha-galactosidase polypeptide useful for treating
 PT lysosomal storage diseases
 PT
 PS Claim 7; Page 51-52; 88pp; English.
 XX
 CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing

CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing
 CC to a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-12R protein.
 XX
 SQ Sequence 423 AA;
 Query Match 100.0%; Score 118; DB 23; Length 423;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TSLRSHINPTGVLLQSEKDEL 24
 Db 400 TSLRSHINPTGVLLQSEKDEL 423
 RESULT 3
 AAE19257
 ID AAE19257 standard; peptide; 28 AA.
 XX
 AC AAE19257;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Human recombinant alpha-galactosidase A modified peptide, rGal-8R.
 XX
 KW Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
 KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
 KW Fabry disease; lysosomal storage disease; human.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200208404-A2.
 XX
 XX 31-JAN-2002.
 PD
 XX 26-JUL-2001; 2001WO-US24111.
 PF
 XX 26-JUL-2000; 2000US-0626127.
 PR
 XX (LARG-) LARGE SCALE BIOLOGY CORP.
 PA
 PI Garger SJ, Turpen TH, Kumagai MH;
 XX
 XX WPI; 2002-195873/25.
 XX
 XX New glucocerebrosidase and alpha-galactosidase having a
 PT post-translational modification, useful in enzyme replacement therapy
 PT for treating lysosomal storage diseases, e.g. Gaucher disease,
 PT Niemann-Pick disease, Fabry disease
 XX
 XX Example 11; Fig 5; 102pp; English.
 XX
 CC The invention relates to production of lysosomal enzymes in plants
 CC by transient expression. The invention particularly relates to
 CC glucocerebrosidase and alpha-galactosidase having a post-translational
 CC modification. The enzymes are useful in enzyme replacement therapy for
 CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease), in researches for
 CC developing new approaches to medical treatment of lysosomal storage
 CC diseases and in industrial processes involving enzymatic substrate
 CC hydrolysis. The present sequence is human recombinant alpha-
 CC galactosidase-A C-terminal modified peptide.
 XX
 SQ Sequence 28 AA;
 Query Match 89.8%; Score 106; DB 23; Length 28;
 Best Local Similarity 85.7%; Pred. No. 4e-10;
 Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

```

QY 1 TSRLRSHINPTGTVLLQL-----SEKDEL 24
DB 1 TSRLRSHINPTGTVLLQLENTMSEKDEL 28

RESULT 4
AAE28211
ID AAE28211 standard; Protein: 427 AA.
AC AAE28211;
XX 27-DEC-2002 (first entry)
XX Human rGAL-8R protein.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-8R.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Cleavage-site 401..402
XX /note= "CTPP cleavage site"
XX US2002088024-A1.
XX 04-JUL-2002.
XX 13-NOV-2001; 2001US-0993059.
XX 26-JUL-2000; 2000US-0626127.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMAGAI/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
XX N-PSDB; AAD45224.
XX Novel human alpha-galactosidase polypeptide useful for treating
XX lysosomal storage diseases
XX Claim 7; Page 46-47; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
XX carboxy terminus and the production of enzymatically active recombinant
XX human and animal lysosomal enzymes. The invention is useful for producing
XX lysosomal enzymes for treating lysosomal storage diseases, producing
XX altered or mutated proteins, enzymatically active or otherwise, to serve
XX as precursors or substrates for further in vivo or in vitro processing
XX to a specialised industrial form for research or therapeutic uses, to
XX produce more effective therapeutic enzyme, for producing antibodies
XX against lysosomal enzymes for medical diagnostic use, and in any
XX commercial process that involves substrate hydrolysis. The present
XX sequence is human rGAL-8R protein.
XX Sequence 427 AA;
Query Match 89.8%; Score:106; DB 23; Length 427;
Best Local Similarity 85.7%; Pred. No. 1e-08;
Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 TSRLRSHINPTGTVLLQL-----SEKDEL 24
DB 400 TSRLRSHINPTGTVLLQLENTMSEKDEL 427

RESULT 5
AAE19255
ID AAE19255 standard; peptide; 32 AA.
XX Query Match 88.1%; Score 104; DB 23; Length 32;
XX Best Local Similarity 75.0%; Pred. No. 9.9e-10;
XX Matches 24; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 TSRLRSHINPTGTVLLQ-----LSEKDEL 24
DB 1 TSRLRSHINPTGTVLLQLENTMQMSLSEKDEL 32

RESULT 6
AAE28209
ID AAE28209 standard; Protein: 431 AA.
XX AAE28209;
XX 27-DEC-2002 (first entry)
XX Human rGAL-4R protein.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-4R.
XX Homo sapiens.
XX Key Location/Qualifiers

```



```

AAE19254
ID AAE19254 standard; peptide; 26 AA.
XX
AC AAE19254;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human recombinant alpha-galactosidase A modified peptide, rGal-4.
XX
KW Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
KW Fabry disease; lysosomal storage disease; human.
XX
OS Homo sapiens.
XX Synthetic.
XX
PN WO200208404-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US24111.
XX
PR 26-JUL-2000; 2000US-0626127.
XX
PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-195873/25.
XX
PF New glucocerebrosidase and alpha-galactosidase having a
PT post-translational modification, useful in enzyme replacement therapy
PT for treating lysosomal storage diseases, e.g. Gaucher disease,
PT Niemann-Pick disease, Fabry disease
XX
PS Example 11; Fig 5; 102pp; English.
XX
CC The invention relates to production of lysosomal enzymes in plants
CC by transient expression. The invention particularly relates to
CC glucocerebrosidase and alpha-galactosidase having a post-translational
CC modification. The enzymes are useful in enzyme replacement therapy for
CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease), in researches for
CC developing new approaches to medical treatment of lysosomal storage
CC diseases and in industrial processes involving enzymatic substrate
CC hydrolysis. The present sequence is human recombinant alpha-
CC galactosidase-A C-terminal modified peptide.
XX
SQ Sequence 26 AA;
Query Match 75.4%; Score 89; DB 23; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGVLLQL 18
DB 1 TSRLRSHINPTGVLLQL 18
RESULT 12
AAE19252
ID AAE19252 standard; peptide; 30 AA.
XX
AC AAE19252;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human recombinant alpha-galactosidase A (rGal-A) peptide.
XX
KW Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
KW Fabry disease; lysosomal storage disease; human.
XX

```

```

OS Homo sapiens.
XX
FH key Location/Qualifiers
FT Cleavage-site 2 /note= "CTPP cleavage site"
FT
XX
PN WO200208404-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US24111.
XX
PR 26-JUL-2000; 2000US-0626127.
XX
PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-195873/25.
XX
PF New glucocerebrosidase and alpha-galactosidase having a
PT post-translational modification, useful in enzyme replacement therapy
PT for treating lysosomal storage diseases, e.g. Gaucher disease,
PT Niemann-Pick disease, Fabry disease
XX
PS Example 11; Fig 5; 102pp; English.
XX
CC The invention relates to production of lysosomal enzymes in plants
CC by transient expression. The invention particularly relates to
CC glucocerebrosidase and alpha-galactosidase having a post-translational
CC modification. The enzymes are useful in enzyme replacement therapy for
CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease), in researches for
CC developing new approaches to medical treatment of lysosomal storage
CC diseases and in industrial processes involving enzymatic substrate
CC hydrolysis. The present sequence is human recombinant alpha-
CC galactosidase-A r(Gal-A) C-terminal peptide.
XX
SQ Sequence 30 AA;
Query Match 75.4%; Score 89; DB 23; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGVLLQL 18
DB 1 TSRLRSHINPTGVLLQL 18
RESULT 13
AAE53771
ID AAE53771 standard; Protein; 92 AA.
XX
AC AAE53771;
XX
DT 25-MAR-2003 (updated)
DT 15-NOV-1994 (first entry)
XX
DE Protein A domain E.
XX
KW Alpha-galactosidase A; Fabry disease; blood group O; blood group B;
KW COS-1; enzyme replacement therapy; glycoconjugate; protein A;
KW IgG binding domain; fusion protein; alpha-Gal A.
XX
OS Staphylococcus aureus.
XX
PN WO9412628-A1.
XX
PD 09-JUN-1994.
XX
PF 30-NOV-1993; 93WO-US11539.
XX
PR 30-NOV-1992; 92US-0983451.

```


XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX Bishop DF, Desnick RJ, Ioannou YA;
 PI WPI: 1994-200257/24.
 DR N-PSDB; AA066242.
 XX Prodn of human alpha-galactosidase A - by culturing cells contg
 PT the coding sequence and the beta-galactosidase alpha-2,6-sialyl
 PT transferase gene and regulation sequences
 XX Disclosure; Page 108; 156pp; English.
 XX The cDNA sequence (AAQ66241) of human alpha-galactosidase A
 CC (alpha-Gal A) was determined. A fusion protein of alpha-Gal A cDNA
 CC and protein A IgG binding domain E (AAQ66242, AAR53771) was expressed
 CC in COS-1 cells. The alpha-Gal A cDNA was amplified by PCR using
 CC primers (AAQ66243-44) designed to delete the termination codon, to
 CC add a 3' collagenase cleavage consensus sequence and to include a
 CC 5' EcoRI recognition site. The protein A domain E was amplified
 CC with the collagenase consensus sequence in the 5' oligonucleotide
 CC using the primers given in AAQ66245-46. The enzyme can be used
 CC for Fabry disease enzyme replacement therapy, to convert blood
 CC group B to O, or to hydrolyze alpha-D-galactosyl residues from
 CC glycoconjugates.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX Query Sequence 92 AA;
 SQ Query Match 75.4%; Score 89; DB 15; Length 92;
 Best Local Similarity 100.0%; Pred. No. 9.9e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TSRLRSHINPTGTVLLQL 18
 Db 3 TSRLRSHINPTGTVLLQL 20
 RESULT 14
 AAB47203
 ID AAB47203 standard; Protein; 386 AA.
 XX AC AAB47203;
 XX 18-JUL-2001 (first entry)
 DE Mutant delta 12 alpha-galactosidase A.
 XX Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant;
 KW deletion; Fabry disease; C-terminal truncation.
 XX Homo sapiens.
 OS US6210666-B1.
 PN 03-APR-2001.
 PD 21-OCT-1998; 98US-0176666.
 PF 21-OCT-1997; 97US-0062560.
 PR (ORPH-) ORPHAN MEDICAL INC.
 PA Miyamura N;
 PI WPI: 2001-289627/30.
 DR N-PSDB; AAC85748.
 XX New alpha-galactosidase A polypeptide, useful for treating Fabry
 PT disease in humans or for treating a condition associated with a
 PT deficiency of alpha-galactosidase A -
 XX

PS Claim 2; Column 47-50; 51pp; English.
 XX This sequence represents a variant human alpha-galactosidase A
 CC (alpha-Gal-A) which has a C-terminal truncation of 12 amino acids.
 CC Alpha-Gal A or its variant, which has a carboxy-terminal deletion of
 CC 2-11 amino acids and which exhibits alpha-galactosidase A enzyme
 CC activity, is useful for treating Fabry disease or a condition associated
 CC with a deficiency of alpha-galactosidase A. Fabry disease has been
 CC linked to an insertion of an adenine nucleotide at nt 1095 in exon 7
 CC of the alpha-Gal A coding sequence, which results in a tyrosine-to-stop
 CC codon substitution at amino acid residue 365 (Y365X), causing a
 CC C-terminal truncation by 65 residues.
 XX Query Sequence 386 AA;
 SQ Query Match 75.4%; Score 89; DB 22; Length 386;
 Best Local Similarity 100.0%; Pred. No. 5.5e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TSRLRSHINPTGTVLLQL 18
 Db 369 TSRLRSHINPTGTVLLQL 386
 RESULT 15
 AAB47202
 ID AAB47202 standard; Protein; 387 AA.
 XX AC AAB47202;
 XX 18-JUL-2001 (first entry)
 DT Mutant delta 11 alpha-galactosidase A.
 DE Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant;
 KW deletion; Fabry disease; C-terminal truncation.
 XX Homo sapiens.
 OS US6210666-B1.
 PN 03-APR-2001.
 PD 21-OCT-1998; 98US-0176666.
 PF 21-OCT-1997; 97US-0062560.
 PR (ORPH-) ORPHAN MEDICAL INC.
 PA Miyamura N;
 PI WPI: 2001-289627/30.
 DR N-PSDB; AAC85747.
 XX New alpha-galactosidase A polypeptide, useful for treating Fabry
 PT disease in humans or for treating a condition associated with a
 PT deficiency of alpha-galactosidase A -
 XX Claim 2; Column 45-48; 51pp; English.
 XX This sequence represents a variant human alpha-galactosidase A
 CC (alpha-Gal-A) which has a C-terminal truncation of 11 amino acids.
 CC Alpha-Gal A or its variant, which has a carboxy-terminal deletion of
 CC 2-11 amino acids and which exhibits alpha-galactosidase A enzyme
 CC activity, is useful for treating Fabry disease or a condition associated
 CC with a deficiency of alpha-galactosidase A. Fabry disease has been
 CC linked to an insertion of an adenine nucleotide at nt 1095 in exon 7
 CC of the alpha-Gal A coding sequence, which results in a tyrosine-to-stop
 CC codon substitution at amino acid residue 365 (Y365X), causing a
 CC C-terminal truncation by 65 residues.
 XX Query Sequence 387 AA;
 SQ

Query Match 75.4%; Score 89; DB 22; Length 387;
Best Local Similarity 100.0%; Pred. No. 5.5e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
|||||
Db 369 TSRLRSHINPTGVLLQL 386

Search completed: July 29, 2003, 16:14:24
Job time : 6.49664 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:09:11 ; Search time 6.65772 Seconds

(without alignments)
930.238 Million cell updates/sec

Title: US-09-993-059-30

Perfect score: 118

Sequence: 1 TSLRSHINPTGTVLLQLSEKDEL 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	46.6	809	16 Q92Z25	Q92Z25 rhizobium m
2	54	45.8	203	16 Q8X8R4	Q8X8R4 escherichia
3	52	44.1	2515	5 Q9W2J8	Q9W2J8 drosophila
4	51	43.2	263	5 Q95T90	Q95T90 drosophila
5	51	43.2	313	5 Q8IH41	Q8IH41 drosophila
6	51	43.2	316	5 Q9I7R1	Q9I7R1 drosophila
7	50	42.4	421	11 Q8BGZ6	Q8BGZ6 mus musculus
8	50	42.4	475	10 Q9ZJ08	Q9ZJ08 arabidopsis
9	49	41.5	248	17 Q8U217	Q8U217 pyrococcus
10	49	41.5	723	10 Q82670	Q82670 cicer ariet
11	49	41.5	1180	16 Q8X6M9	Q8X6M9 escherichia
12	48.5	41.1	188	12 Q90386	Q90386 rhesus cyto
13	48	40.7	215	16 Q8P5S1	Q8P5S1 xanthomonas
14	48	40.7	1183	16 Q8FEB3	Q8FEB3 escherichia
15	47	39.8	348	12 Q8V9U4	Q8V9U4 african swi
16	47	39.8	417	11 Q70629	Q70629 mus musculus

17	47	39.8	417	11	Q88691	Q88691 mus musculus
18	47	39.8	568	16	Q92626	Q92626 listeria in
19	47	39.8	962	12	Q65162	Q65162 african swi
20	47	39.8	1731	5	Q9TXP3	Q9TXP3 caenorhabdi
21	46.5	39.4	209	17	Q8PZK0	Q8PZK0 methanosarc
22	46	39.0	257	17	Q30121	Q30121 archaeoglob
23	46	39.0	499	16	Q07268	Q07268 mycobacteri
24	46	39.0	609	5	Q22680	Q22680 caenorhabdi
25	46	39.0	643	16	Q8VKD9	Q8VKD9 mycobacteri
26	46	39.0	843	10	Q93X58	Q93X58 fragaria an
27	46	39.0	847	2	Q8KX4	Q8KX4 rhizobium l
28	46	39.0	1249	11	Q9QXV2	Q9QXV2 mus musculu
29	46	39.0	1249	11	Q920Q2	Q920Q2 mus musculu
30	46	39.0	4135	6	Q18977	Q18977 bos taurus
31	45	38.1	150	5	Q9W383	Q9W383 drosophila
32	45	38.1	240	16	Q8CMH3	Q8CMH3 streptococc
33	45	38.1	264	17	Q26764	Q26764 methanobact
34	45	38.1	354	15	Q9IU69	Q9IU69 human immun
35	45	38.1	512	16	Q8CV14	Q8CV14 oceanobacil
36	45	38.1	539	10	Q80914	Q80914 arabidopsis
37	45	38.1	774	4	Q95941	Q95941 homo sapien
38	45	38.1	899	10	Q9SG75	Q9SG75 arabidopsis
39	45	38.1	917	10	Q9CAF6	Q9CAF6 arabidopsis
40	45	38.1	1152	16	Q985U4	Q985U4 rhizobium l
41	45	38.1	1250	4	Q9C0J4	Q9C0J4 homo sapien
42	45	38.1	1251	4	Q9UB29	Q9UB29 homo sapien
43	45	38.1	15281	3	Q09164	Q09164 tolypocladi
44	44.5	37.7	214	3	Q9UTZ6	Q9UTZ6 schizosacch
45	44.5	37.7	438	5	Q8T071	Q8T071 drosophila

ALIGNMENTS

RESULT 1

Q92Z25 PRELIMINARY; PRT; 809 AA.
 ID Q92Z25
 AC Q92Z25;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Adenylate cyclase, putative.
 GN RA0303 OR SMA0579.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021.
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Guzy J.,
 RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymA megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL; AF007222; AAK64961.1; -;
 DR InterPro; IPR001054; G_cyclase.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00211; guanylate_cyc; 1.
 DR Pfam; PF00672; HAMP; 1.
 DR SMART; SM00044; CYCC; 1.
 DR SMART; SM00304; HAMP; 1.
 DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 809 AA; 88949 MW; F650FF4BD67BA21 CRC64;
 Query Match 46.6%; Score 55; DB 16; Length 809;
 Best Local Similarity 44.4%; Pred. No. 4.2;

10

```

OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY060280; AAL25319.1; -
DR FlyBase: FBgn0036211; CG5946
DR InterPro: IPR001834; Cyt_B5_reductase.
DR InterPro: IPR001709; FPN_cyt_redctse.
DR InterPro: IPR001433; Oxred_FAD/NAD(P).
DR Pfam: PF00970; FAD_binding_6; 1..
DR Pfam: PF00175; NAD_binding_1; 1.
DR PRINTS: PR00406; CYTB5RDTASE.
DR PRINTS: PR00371; FPNCR.
SQ SEQUENCE 263 AA; 29654 MW; 6F76B0464FE21526 CRC64;

Query Match 43.2%; Score 51; DB 5; Length 263;
Best Local Similarity 45.8%; Pred. No. 5.6;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQSEKDEL 24
   I:||||:|:| | | | | |
Db 44 TARLRLTVDPNKYLPLIERENL 67

RESULT 5
Q8IH41 PRELIMINARY; PRT; 313 AA.
ID Q8IH41;
AC Q8IH41;
DT 01-WAR-2003 (Tremblrel. 23, Created)
DT 01-WAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-WAR-2003 (Tremblrel. 23, Last annotation update)
DE GH26062p.
GN CG5946.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=y;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BT001444; AAN71199.1; -
DR EMBL: BT001444; AAN71199.1; -
SQ SEQUENCE 313 AA; 35296 MW; 7FAF7D07478A7850 CRC64;

Query Match 43.2%; Score 51; DB 5; Length 313;
Best Local Similarity 45.8%; Pred. No. 6.7;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQSEKDEL 24
   I:||||:|:| | | | | |
Db 41 TARLRLTVDPNKYLPLIERENL 64

RESULT 6
Q9I7R1 PRELIMINARY; PRT; 316 AA.
ID Q9I7R1
AC Q9I7R1; Q9VTP2;
DT 01-WAR-2001 (Tremblrel. 16, Created)
DT 01-WAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-WAR-2001 (Tremblrel. 16, Last annotation update)
DT 01-WAR-2003 (Tremblrel. 23, Last annotation update)

```

```

DE GN Drosophila melanogaster (Fruit fly).
OS CG5946.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
G "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- CATALYTIC ACTIVITY: NADH + 2 FERRICYTOCHROME B5 = NAD(+) + 2
CC -!- FERROCYTOCHROME B5.
CC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC -!- SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL: AE003543; AAF50004.1; -
DR EMBL: AE003543; AAG22320.1; -
DR HSSP: P17571; 2CND.
DR FlyBase: FBgn0036211; CG5946.
DR InterPro: IPR001834; Cyt_B5_reductase.
DR InterPro: IPR001709; FPN_cyt_redctse.
DR InterPro: IPR001433; Oxred_FAD/NAD(P).
DR InterPro: IPR001221; Phe_hydroxylase.
DR Pfam: PF00970; FAD_binding_6; 1.
DR Pfam: PF00175; NAD_binding_1; 1.
DR PRINTS: PR00406; CYTB5RDTASE.
DR PRINTS: PR00371; FPNCR.
DR PRINTS: PR00410; PHEHYDRXLASE.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Hypothetical protein;
FT NP_BIND 146 176 FAD (NAD PART) (BY SIMILARITY).
FT NP_BIND 185 221 FAD (NAD PART) (BY SIMILARITY).
FT VARSPPLIC 1 8 MARIYESN -> MTEFD (IN SHORT ISOFORM).
SQ SEQUENCE 316 AA; 35573 MW; 8F872420A7776E0F CRC64;

```

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.


```

RC STRAIN-ATCC 33913 / NCPPB 528;
RX MEDLINE-2202145; PubMed-12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Canavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AE012443; AAM42535.1; -.
DR InterPro; IPR001034; HTH_Deor.
DR Pfam; PF004788; RpiA.
DR ProDom; PD005813; RpiA; 1.
DR TIGRFAMs; TIGR00021; rpiA; 1.
KW Isomerase; Complete proteome.
SQ SEQUENCE 215 AA; 22919 MW; 3B592FB1D9213E07 CRC64;

Query Match 40.7%; Score 48; DB 16; Length 215;
Best Local Similarity 41.4%; Pred. No. 14;
Matches 12; Conservative 4; Mismatches 7; Indels 6; Gaps 1;

Qy 1 TSLRSH- - - - - INPTGTVLLQSEKDE 23
Db 55 TARLRQHGVIEVDLNLHHTGTLSTLYVDGADE 83

RESULT 14
Q8FEB3 PRELIMINARY; PRT; 1183 AA.
AC Q8FEB3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN RECB OR C3414.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE-22398234; PubMed-12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016765; AAN81859.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 1183 AA; 134423 MW; F1LCB6919C16B8CA CRC64;

Query Match 40.7%; Score 48; DB 16; Length 1183;
Best Local Similarity 43.5%; Pred. No. 88;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQSEKDE 23
Db 998 TAVLQAPLNETGVSQNLSDRDK 1020

```

```

RESULT 15
Q8V9U4 PRELIMINARY; PRT; 348 AA.
AC Q8V9U4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE L09AR protein.
GN L09AR.
OS African swine fever virus (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10497;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Malawi Lil-20/1;
RA Roberts P.C., Lu Z., Rock D.L.;
RT "Nucleotide sequence and analysis of 16.25 kilobase pairs of the
RT African swine fever virus genome that span the central variable
RT region.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; L00966; AAL31318.1; -.
SQ SEQUENCE 348 AA; 40205 MW; B65835AADC4A4A0B CRC64;

Query Match 39.8%; Score 47; DB 12; Length 348;
Best Local Similarity 39.1%; Pred. No. 34;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQSEKDE 23
Db 88 TSLRKEMVNEPQDVTARELNQKE 110

Search completed: July 29, 2003, 16:17:13
Job time : 8.65772 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 15:42:09 ; Search time 1.50336 Seconds
(without alignments)

750.748 Million cell updates/sec

Title: US-09-993-059-30

Perfect score: 118

Sequence: 1 TSLRLSHINPTGTVLLQLSEKDEL 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	89	75.4	429	1	AGAL_HUMAN	P06280	homo sapien
2	54	45.8	162	1	SEEB_ECOLI	P38392	escherichia
3	52	44.1	2515	1	TUD_DROME	P25823	drosophila
4	50.5	42.8	82	1	VR3_BPT4	P17309	bacterioph
5	50	42.4	419	1	AGAL_MOUSE	P51569	mus musculu
6	50	42.4	1065	1	CARB_BACCL	P46537	bacillus ca
7	48	40.7	196	1	Y009_NEIMB	Q91xs2	neisseria m
8	48	40.7	215	1	RPIA_XANCP	Q8p5s1	xanthomonas
9	47	39.8	417	1	PEDF_MOUSE	P97298	mus musculu
10	46	39.0	278	1	PYRF_PENCH	P09463	penicillium
11	45.5	38.6	557	1	UVRC_THENA	Q9wyv3	thermotoga
12	45	38.1	300	1	SPSY_YEAST	Q12455	saccharomyc
13	45	38.1	542	1	RESB_BACSU	P35161	bacillus su
14	45	38.1	1071	1	CARB_NEIGO	Q59599	neisseria g
15	45	38.1	1071	1	CARB_NEIMA	Q91w02	neisseria m
16	45	38.1	1071	1	CARB_NEIMB	Q91xw8	neisseria m
17	44.5	37.7	464	1	EBFL_DROME	Q9vph7	drosophila
18	44.5	37.7	522	1	MG51_SCHEFO	P87323	schizosacch
19	44	37.3	285	1	Y091_MYCTU	Q10551	mycobacteri
20	44	37.3	288	1	ISPE_CLOPE	Q8xia9	clostridium
21	44	37.3	451	1	PKR6_HUMAN	Q13882	homo sapien
22	44	37.3	579	1	PHOR_BACSU	P23545	bacillus su
23	44	37.3	2240	1	C089_DROME	Q9veul	drosophila
24	43	36.4	215	1	RPIA_XANAC	Q8p4n9	xanthomonas
25	43	36.4	267	1	PYRF_KLUOLA	P07922	kluiveromyc
26	43	36.4	286	1	AAC2_ACTIB	P29807	acinetobact
27	43	36.4	286	1	AAC2_SALSP	P13246	salmonella
28	43	36.4	286	1	AAC3_SALSP	P13245	salmonella
29	43	36.4	296	1	Y608_TREPA	O83617	treponema p
30	43	36.4	707	1	J1P1_MOUSE	Q9wv19	mus musculu
31	43	36.4	708	1	J1P1_RAT	Q9w237	r c-jun-ami
32	43	36.4	711	1	J1P1_HUMAN	Q9uqf2	homo sapien
33	42.5	36.0	256	1	VBRI_PHUV	Q06927	pepper huas

RESULT 1

ID	AGAL_HUMAN	STANDARD;	PRT;	429 AA.
AC	P06280;			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	01-AUG-1988	(Rel. 08, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase alfa).			
DE	galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase alfa).			
GN	GLA.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]	SEQUENCE FROM N.A.		
RP	TISSUE=Lymphoblast;			
RC	TISSUE=Lymphoblast;			
RX	MEDLINE=89263745; PubMed=2542896;			
RA	Kornreich R., Desnick R.J., Bishop D.F.;			
RT	"Nucleotide sequence of the human alpha-galactosidase A gene.";			
RL	Nucleic Acids Res. 17:3301-3302(1989).			
RN	[2]	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fibroblast;			
RX	MEDLINE=87246603; PubMed=3036505;			
RA	Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,			
RA	Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,			
RA	Ginns E.I.;			
RT	"Signal sequence and DNA-mediated expression of human lysosomal			
RL	alpha-galactosidase A.";			
RN	[3]	Eur. J. Biochem. 165:275-280(1987).		
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95352959; PubMed=7626884;			
RA	Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,			
RA	Gibbs R.A.;			
RT	"Sixty-nine kilobases of contiguous human genomic sequence containing			
RL	the alpha-galactosidase A and Bruton's tyrosine kinase loci.";			
RN	[4]	Mamm. Genome 6:334-338(1995).		
RP	SEQUENCE FROM N.A.			
RX	SEQUENCE FROM N.A.			
RA	Wilson S.;			
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.			
RN	[5]	SEQUENCE FROM N.A.		
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Uterus;			
RX	MEDLINE=23388957; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,			
RA	Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.F., Bhat N.K.,			
RA	Altshul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Hopkins R.F., Jordan H., Jordan K., Farmer A., Rubin G.M., Hong L.,			
RA	Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			

34	42.5	36.0	1454	1	CPSA_CAEEL	Q9n4c2	caenorhabdi
35	42	35.6	95	1	FER4_RHOCA	P16022	rhodobacter
36	42	35.6	125	1	GDF6_MOUSE	P43028	mus musculu
37	42	35.6	151	1	GDF7_MOUSE	P43029	mus musculu
38	42	35.6	243	1	TPX1_MOUSE	P15563	mus musculu
39	42	35.6	271	1	POBR_ACICA	Q43992	acinetobact
40	42	35.6	421	1	AMYA_VIGMU	P17859	vigna mungo
41	42	35.6	423	1	PHS2_PHALU	Q43617	phaseolus l
42	42	35.6	428	1	PHS1_PHALU	P80463	phaseolus l
43	42	35.6	436	1	GDF6_BOVIN	P55106	bos taurus
44	42	35.6	1035	1	DPOL_RHCM6	O71121	rhesus cyto
45	42	35.6	1064	1	CARB_BACST	O50302	bacillus st

ALIGNMENTS

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalley D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL *Proc. Natl. Acad. Sci. U.S.A.* 99:16899-16903(2002).
RN [16]
RN SEQUENCE OF 31-429 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Desnick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL *Proc. Natl. Acad. Sci. U.S.A.* 83:4859-4863(1986).
RN [7]
RN SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE=88112869; PubMed=2892762;
RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
RT human lysosomal enzyme, alpha-galactosidase A.";
RL *Gene* 58:177-188(1987).
RN [8]
RN SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE=88234528; PubMed=2836863;
RA Bishop D.F., Kornreich R., Desnick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
RT further evidence for the absence of a 3' untranslated region.";
RL *Proc. Natl. Acad. Sci. U.S.A.* 85:3903-3907(1988).
RN [9]
RN RNA EDITING.
RX MEDLINE=95380278; PubMed=7503918;
RA Novo F.J., Kruszewski A., Mcdermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine
RT to purine conversion.";
RL *Nucleic Acids Res.* 23:2636-2640(1995).
RN [10]
RN REVIEW ON FD VARIANTS.
RX MEDLINE=94258158; PubMed=7911050;
RA Eng C.M., Desnick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
RT human alpha-galactosidase A gene.";
RL *Hum. Mutat.* 3:103-111(1994).
RN [11]
RN VARIANT FD SER-40;
RX MEDLINE=90092580; PubMed=2152885;
RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
RA Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
RT activity caused by a single amino acid substitution of Pro-40 by
RT Ser.";
RL *FEBS Lett.* 259:353-356(1990).
RN [12]
RN VARIANT FD VAL-296.
RX MEDLINE=91101674; PubMed=1846223;
RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
RA Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
RT to the myocardium.";
RL *New Engl. J. Med.* 324:395-399(1991).
RN [13]
RN VARIANT FD GLN-301.
RX MEDLINE=91022721; PubMed=2171331;
RA Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
RA Bishop D.F., Desnick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
RT in classical and atypical hemizygotes with Fabry disease.";
RN [14]

Am. J. Hum. Genet. 47:784-789(1990).
RN [14]
RN VARIANT FD TRP-356.
RX MEDLINE=89198098; PubMed=2539398;
RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
RA Sakuraba H., Desnick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
RT in the alpha-galactosidase gene.";
RL *J. Clin. Invest.* 83:1390-1399(1989).
RN [15]
RN VARIANTS FD GLN-66; CYS-112; GLU-279; GLN-301 AND ARG-328.
RX MEDLINE=92250012; PubMed=1315715;
RA Ishii S., Sakuraba H., Suzuki Y.;
RT "Point mutations in the upstream region of the alpha-galactosidase A
RT gene exon 6 in an atypical variant of Fabry disease.";
RL *Hum. Genet.* 89:29-32(1992).
RN [16]
RN VARIANTS SER-34; GLY-56; ARG-162; GLN-227; VAL-264; PHE-297;
RX TYR-313; ALA-328 AND ARG-404 DEL.
RX MEDLINE=94070864; PubMed=7504405;
RA Eng C.M., Resnick-Silverman L.A., Niehaus D.J., Astrin K.H.,
RA Desnick R.J.;
RT "Nature and frequency of mutations in the alpha-galactosidase A gene
RT that cause Fabry disease.";
RL *Am. J. Hum. Genet.* 53:1186-1197(1993).
RN [17]
RN VARIANTS FD SER-34; SER-215; ALA-269; LYS-327 AND ARG-361.
RX MEDLINE=93372810; PubMed=8395937;
RA Davies J.P., Winchester B.G., Malcolm S.;
RT "Mutation analysis in patients with the typical form of
RT Anderson-Fabry disease.";
RL *Hum. Mol. Genet.* 2:1051-1053(1993).
RN [18]
RN VARIANTS FD ARG-35; LEU-49; VAL-165 AND GLU-316.
RX MEDLINE=94348524; PubMed=8069316;
RA Davies J.P., Christomanou H., Winchester B.G., Malcolm S.;
RT "Detection of 8 new mutations in the alpha-galactosidase A gene in
RT Fabry disease.";
RL *Hum. Mol. Genet.* 3:667-669(1994).
RN [19]
RN VARIANTS FD.
RX MEDLINE=95152501; PubMed=7531540;
RA Eng C.M., Niehaus D.J., Enriquez A.L., BURGERT T.S., Ludman M.D.,
RA Desnick R.J.;
RT "Fabry disease: twenty-three mutations including sense and antisense
RT Cpg alterations and identification of a deletion hot-spot in the
RT alpha-galactosidase A gene.";
RL *Hum. Mol. Genet.* 3:1795-1799(1994).
RN [20]
RN VARIANTS FD GLN-66; CYS-112; VAL-156; VAL-166; ALA-260; GLU-279;
RX ILE-296; GLN-301; LYS-320; ARG-328 AND SER-373.
RX MEDLINE=96024628; PubMed=7575533;
RA Okumura T., Ishii S., Takenaka T., Kase R., Kamei S., Sakuraba H.,
RA Suzuki Y.;
RT "Galactose stabilizes various missense mutants of alpha-galactosidase
RT in Fabry disease.";
RL *Biochem. Biophys. Res. Commun.* 214:1219-1224(1995).
RN [21]
RN VARIANTS FD TYR-142; VAL-156 AND VAL-166.
RX MEDLINE=95278912; PubMed=7759078;
RA Okumura T., Ishii S., Kase R., Kamei S., Sakuraba H., Suzuki Y.;
RT "Alpha-galactosidase gene mutations in Fabry disease: heterogeneous
RT expressions of mutant enzyme proteins.";
RL *Hum. Genet.* 95:557-561(1995).
RN [22]
RN VARIANTS FD PRO-32; SER-34; ASP-85; THR-156 AND GLN-301.
RX MEDLINE=95322989; PubMed=7599642;
RA Madsen K.M., Hasholt L., Soerensen S.A., Lagerstroem Fermer M.,
RA Dahl N.;
RT "Two novel mutations (L32P) and (G85N) among five different missense
RT mutations in six Danish families with Fabry's disease.";
RL *Hum. Mutat.* 5:277-278(1995).
RN [23]

Query Match 75.4%; Score 89; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
DB 400 TSRLRSHINPTGTVLLQL 417

RESULT 2

STBL_ECOLI STANDARD; PRT; 162 AA.
AC P38392; P76059; P76850;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Superinfection exclusion protein B.
GN SIEB OR B1353.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sanpei K., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 1-75 FROM N.A.
RC STRAIN=B;
RX MEDLINE=94148774; PubMed=7508908;
RA Faubladier M., Bouche J.-P.;
RT "Division inhibition gene dicF of Escherichia coli reveals a
widespread group of prophage sequences in bacterial genomes.";
RL J. Bacteriol. 176:1150-1156(1994).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AE000233; AAC74435.1; ALT_INIT.
DR EMBL; D90774; BAA14956.1; -
DR EMBL; Z23096; CAA80642.1; -
DR Ecogene; EG12154; sieB.
KW Transmembrane; Complete proteome.
FT TRANSMEM 48 POTENTIAL.
SQ SEQUENCE 162 AA; 19244 MW; 8DC2A2509E315955 CRC64;

Query Match 45.8%; Score 54; DB 1; Length 162;
Best Local Similarity 51.9%; Pred. No. 0.38;

Matches 14; Conservative 3; Mismatches 4; Indels 6; Gaps 1;
QY 2 SRLRSHIN-----PTGTVLLQLSEKD 22
DB 68 SRLRAHQKHYSLLPEQRVLLRLSEKE 94

RESULT 3

TUD_DROME STANDARD; PRT; 2515 AA.
AC P25823;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maternal tudor protein.
GN TUD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92038995; PubMed=1936993;
RA Golumbeski G.S., Bardsley A., Tax F., Boswell R.E.;
RT "Tudor, a posterior-group gene of Drosophila melanogaster, encodes a
novel protein and an mRNA localized during mid-oogenesis.";
RL Genes Dev. 5:2060-2070(1991).
CC -!- FUNCTION: REQUIRED DURING OOGENESIS FOR THE FORMATION OF
PRIMORDIAL GERM CELLS AND FOR NORMAL ABDOMINAL SEGMENTATION.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE.
CC -!- MISCELLANEOUS: THE TUD MRNA ACCUMULATES WITHIN THE POSTERIOR
REGION OF THE DEVELOPING OOCYTE DURING THE EARLY TO MIDDLE STAGES
OF OOGENESIS.
CC -!- SIMILARITY: Contains 9 Tudor domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X62420; CAA44286.1; -
DR PIR; A41519; A41519.
DR HSSP; Q16637; 1G5V.
DR Flybase; FBgn0003891; tud.
DR GO; GO:0019090; P:mitochondrial rRNA, mitochondrial export; IMP.
DR GO; GO:0007315; P:pole plasm assembly; IMP.
DR InterPro; IPR001097; Maternal_tudor.
DR InterPro; IPR002999; Tudor.
DR Pfam; PF00567; TUDOR; 10.
DR SMART; SM00333; TUDOR; 10.
DR PROSITE; PS50304; TUDOR; 9.
DR Developmental protein; Repeat.
FT DOMAIN 455 513
FT DOMAIN 641 696
FT DOMAIN 1062 1122
FT DOMAIN 1355 1414
FT DOMAIN 1662 1718
FT DOMAIN 1839 1898
FT DOMAIN 2023 2082
FT DOMAIN 2211 2269
FT DOMAIN 2392 2451
SQ SEQUENCE 2515 AA; 285236 MW; 683C100AD308BADA CRC64;

Query Match 44.1%; Score 52; DB 1; Length 2515;
Best Local Similarity 41.7%; Pred. No. 15;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQLSEKDEL 24
| : : : | : | | | | |

```

Db      1981 TKAIITHVENTSRIYLQFSEKDSL 2004
RESULT 4
VR3_BPT4
ID      VR3_BPT4      STANDARD;      PRT;      82 AA.
AC      P17309;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-OCT-1993 (Rel. 27, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Protein RIII.
GN      RIII OR 31.-1.
OS      Bacteriophage T4.
OC      Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC      T4-like viruses.
OX      NCBI_TaxID=10665;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=D;
RX      MEDLINE=90301484; PubMed=2362813;
RA      Prilipov A.G., Mesyanzhinov V.V., Aebi U., Kellenberger E.;
RT      "Cloning and sequencing of bacteriophage T4 genes between map
RT      positions 128.3-130.3.";
RL      Nucleic Acids Res 18:3635-3635(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=90332452; PubMed=2377483;
RA      Raudonikiene A., Nivinskas R.;
RT      "Nucleotide sequence of bacteriophage T4 gene 31 region.";
RL      Nucleic Acids Res 18:4280-4280(1990).
RN      [3]
RP      SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC      MEDLINE=92267389; PubMed=1587487;
RA      Raudonikiene A., Nivinskas R.;
RT      "Gene rIII is the nearest downstream neighbour of bacteriophage T4
RT      gene 31.";
RL      Gene 114:85-90(1992).
RN      [4]
RP      SEQUENCE FROM N.A., AND MUTANTS.
RC      MEDLINE=94063508; PubMed=8244025;
RA      Raudonikiene A., Nivinskas R.;
RT      "The sequences of gene rIII of bacteriophage T4 and its mutants.";
RN      [5]
RP      SEQUENCE FROM N.A.
RC      Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosis G.,
RA      Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT      "Bacteriophage T4 genome analysis.";
RL      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X17657; CAA35652.1; -
DR      EMBL; M37882; AAA32507.1; -
DR      EMBL; X54536; CAA38406.1; -
DR      EMBL; AF158101; AAD42650.1; -
DR      PIR; JT0765; JT0765.
FT      VARIANT      42      42      H -> R (IN R67 MUTANT; R-TYPE PLAQUE
FT      MORPHOLOGY).
FT      VARIANT      82      82      K -> E (IN RES40 AND RBB9 MUTANTS; R-TYPE
FT      PLAQUE MORPHOLOGY).
FT      CONFLICT      64      82      KPLASAAARKAVRHVVTLK -> NLWLLLHEKQFVTLW
FT      (IN REF. 1).
SQ      SEQUENCE      82 AA; 9325 MW;
Query Match      42.88; Score 50.5; DB 1; Length 82;
Best Local Similarity      52.28; Pred. No. 0.62;
Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;
QY      3 RLRSHPNTGTVL-LQLSEKDEL 24
      | :||| | :|||
Db      38 RYFKHLNPAQTALAAELQEKDEL 60
      | :||| | :|||
RESULT 5
AGAL_MOUSE
ID      AGAL_MOUSE      STANDARD;      PRT;      419 AA.
AC      P51569;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE      galactoside galactohydrolase) (Alpha-D-galactosidase A).
GN      GLA OR AGS.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Kidney;
RX      MEDLINE=96125203; PubMed=8543175;
RA      Ohshima T., Murray G.J., Nagle J.W., Quirk J.M., Kraus M.H.,
RA      Barton N.W., Brady R.O., Kulkarni A.B.;
RT      "Structural organization and expression of the mouse gene encoding
RT      alpha-galactosidase A.";
RL      Gene 166:277-280(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C129;
RX      MEDLINE=95352959; PubMed=7626884;
RA      Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,
RA      Gibbs R.A.;
RT      "Sixty-nine kilobases of contiguous human genomic sequence containing
RT      the alpha-galactosidase A and Bruton's tyrosine Kinase loci.";
RL      Mamm. Genome 6:334-338(1995).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=96316016; PubMed=8733892;
RA      Gotlib R.W., Bishop D.F., Wang A.M., Zeidner K.M., Ioannou Y.I.,
RA      Adler D.A., Distche C.M., Desnick R.J.;
RT      "The entire genomic sequence and cDNA expression of mouse alpha-
RT      galactosidase A.";
RL      Biochem. Mol. Med. 57:139-148(1996).
CC      -1- CATALYTIC ACTIVITY: Melibiose + H(2)O = galactose + glucose.
CC      -1- SUBCELLULAR LOCATION: Lysosomal.
CC      -1- SIMILARITY: BELONGS TO FAMILY 27 OF GLYCOSYL HYDROLASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U34071; AAA96749.1; -
DR      EMBL; L46651; AAA74453.1; -
DR      EMBL; U58105; AAB47244.1; -
DR      EMBL; U50716; AAC52584.1; -
DR      EMBL; U50715; AAC52583.1; -
DR      PIR; JC4522; JC4522.
DR      MGD; MGI:1347344; Gla.
DR      InterPro; IPR002241; Glyco_hydro_27.
DR      InterPro; IPR000111; Glyco_hydro_GHD.
DR      Pfam; PF02065; Melibiase; 1.
DR      PRINTS; PR00740; GLHYDLASE27.
DR      PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR      PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW      Hydrolase; Glycosidase; Glycoprotein; Lysosome; Signal.

```

```

FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT FT 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT FT 930 1065 ALLOSTERIC DOMAIN.
FT FT 1 546
FT REPEAT 547 1065
FT FT NP_BIND 153 210 ATP (POTENTIAL).
FT FT NP_BIND 302 352 ATP (POTENTIAL).
FT FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1065 AA; 116482.MW; 4FEA3A94BBDB4B1 CRC64;

Query Match 42.4%; Score 50; DB 1; Length 1065;
Best Local Similarity 52.9%; Pred. No. 12;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 7 HINPTGVLLQSEKDE 23
||| ||| :|||
Db 931 HIQPHGAVLLTVADKDK 947

RESULT 7
ID YJ09_NEIMB STANDARD; PRT; 196 AA.
AC Q9JXS2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maf-like protein NMB1909.
GN NMB1909.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OC NCBI_TaxID=491;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; Pubmed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittono H., Clark E.B.,
RA Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamachavan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE MAF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002540; AAF42239.1; -
DR PIR: D81028; D81028.
DR TIGR: NMB1909; -
DR HAMAP: MF_00528; -; 1.
DR InterPro: IPR003697; Maf.
DR Pfam: PF02545; Maf; 1.
DR TIGRFAMs: TIGR00172; maf; 1.
DR Complete proteome.
KW ACT_SITE 36
SQ SEQUENCE 196 AA; 21911 MW; C545A168978561B9 CRC64;
POTENTIAL.

```



```
RESULT 14
CARB_NEIGO STANDARD; PRT; 1071 AA.
AC Q59599;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB;
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH811;
RX MEDLINE=95291461; PubMed=7773412;
RA Lawson F.S., Billows F.M., Dillon J.A.;
RT "Organization of carbamoyl-phosphate synthase genes in Neisseria
gonorrhoeae includes a large, variable intergenic sequence which is
also present in other Neisseria species.";
RL Microbiology 141:1183-1191(1995).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Coudert E.;
RL Unpublished observations (JUL-2002).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate. (By similarity).
CC -!- COFACTOR: Binds 3 manganese ions per subunit.
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate.
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -!- CAUTION: This is a conceptual translation; fourteen frameshifts
were introduced in the sequence to maximize the similarity with
other orthologs.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U11295; AAA74996.1; ALT_FRAME.
CC HSSP: P00968; 1A9X.
CC HAMAP: MF_01210; -. 1.
CC InterPro: IPR006275; CarA_L_glu.
CC InterPro: IPR005483; CPhase_L.
CC InterPro: IPR005479; CPhase_L_D2.
CC InterPro: IPR005480; CPhase_L_D3.
CC InterPro: IPR005481; CPhase_L_N.
CC InterPro: IPR004362; MGS like.
CC Pfam: PF00289; CPhase_L_chain; 2.
CC Pfam: PF02786; CPhase_L_D2; 2.
CC Pfam: PF02787; CPhase_L_D3; 1.
CC Pfam: PF02142; MGS; 1.
CC PRINTS: PR00098; CPhase.
CC TIGRfams: TIGR01369; CPhaseI_lrg; 1.
CC PROSITE: PS00866; CPhase_1; 1.
CC PROSITE: PS00867; CPhase_2; 2.
CC ARGININE BIOSYNTHESIS; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese.
FT DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 404 548 OLIGOMERIZATION DOMAIN.
FT DOMAIN 549 930 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 931 1071 ALLOSTERIC DOMAIN.
FT REPEAT 1 548
FT REPEAT 549 1071
```

```
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 303 354 ATP (POTENTIAL).
FT METAL 285 285 MANGANESE 1 (BY SIMILARITY).
FT METAL 299 299 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 2 (BY SIMILARITY).
FT METAL 823 823 MANGANESE 3 (BY SIMILARITY).
FT METAL 835 835 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1071 AA; 117582 MW; 04B99981D3EB8212 CRC64;
Query Match 38.1%; Score 45; DB 1; Length 1071;
Best Local Similarity 43.8%; Pred. No. 72;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Qy 8 INPTGTVLLQLSEKDE 23
Db 933 LNPTGKIFLSVREEDK 948
RESULT 15
CARB_NEIMA STANDARD; PRT; 1071 AA.
AC Q9JW02;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB OR NNA0602
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.B., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491.";
RL Nature 404:502-506(2000).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
CC -!- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL162753; CAB83892.1; -.
CC PIR: F81979; F81979.
CC HSSP: P00968; 1A9X.
CC HAMAP: MF_01210; -. 1.
CC InterPro: IPR006275; CarA_L_glu.
CC InterPro: IPR005483; CPhase_L.
CC InterPro: IPR005479; CPhase_L_D2.
CC InterPro: IPR005480; CPhase_L_D3.
CC InterPro: IPR005481; CPhase_L_N.
CC PROSITE: PS00866; CPhase_1; 1.
CC PROSITE: PS00867; CPhase_2; 2.
CC ARGININE BIOSYNTHESIS; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese.
FT DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 404 548 OLIGOMERIZATION DOMAIN.
FT DOMAIN 549 930 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 931 1071 ALLOSTERIC DOMAIN.
FT REPEAT 1 548
FT REPEAT 549 1071
```

```

DR InterPro; IPR004362; MGS_like.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 1.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
ET DOMAIN 1 403 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
ET DOMAIN 404 548 OLIGOMERIZATION DOMAIN.
ET DOMAIN 549 930 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
ET DOMAIN 931 1071 ALLOSTERIC DOMAIN.
ET REPEAT 1 548
ET REPEAT 549 1071
ET NP_BIND 153 210 ATP (POTENTIAL).
ET NP_BIND 303 354 ATP (POTENTIAL).
ET METAL 285 285 MANGANESE 1 (BY SIMILARITY).
ET METAL 299 299 MANGANESE 1 AND 2 (BY SIMILARITY).
ET METAL 301 301 MANGANESE 2 (BY SIMILARITY).
ET METAL 823 823 MANGANESE 3 (BY SIMILARITY).
ET METAL 835 835 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1071 AA; 117419 MW; 73B39CBD06729974 CRC64;

Query Match      38.1%; Score 45; DB 1; Length 1071;
Best Local Similarity 43.8%; Pred. No. 72;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy      8 INPTGTVLLQLSEKDE 23
       :| | | | : | : | :
Db      933 LNPTGKIFLSVREEDK 948

```

Search completed: July 29, 2003, 16:15:01
 Job time : 2.50336 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:12:30 ; Search time 2.68456 Seconds
(without alignments)
859.748 Million cell updates/sec

Title: US-09-993-059-30

Perfect score: 118

Sequence: 1 TSLRSHINPTGTVLLQLSEKDEL 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	75.4	429	1 GBHUA	alpha-galactosidas
2	55	46.6	809	2 G95299	probable adenylate
3	54	45.8	203	2 D64885	superinfection exc
4	54	45.8	203	2 A90871	phage superinfecti
5	54	45.8	203	2 H85747	superinfection exc
6	52	44.1	2515	2 A41519	posterior-group pr
7	50.5	42.8	82	2 JT0765	rapid lysis III pr
8	50	42.4	419	2 JC4522	alpha-galactosidas
9	50	42.4	475	2 A86372	53.0K hypothetical
10	50	42.4	1065	1 A40169	carbamoyl-phosphat
11	49	41.5	1180	2 E91088	DNA helicase RecB
12	49	41.5	1180	2 G85933	DNA helicase RecB
13	48	40.7	196	2 H81028	Maf/Ycef/Vhde fami
14	47	39.8	568	2 AG1793	para-aminobenzoate
15	46	39.0	257	2 C69264	conserved hypothet
16	46	39.0	275	1 DCFLOC	orotidine-5'-phosp
17	46	39.0	499	2 A70822	probable lpdB prot
18	46	39.0	609	2 T25120	hypothetical prote
19	46	39.0	4135	2 T42629	tenascin-X - bovin
20	45.5	38.6	557	2 H72400	excinuclease ABC C
21	45	38.1	264	2 H65188	hypothetical prote
22	45	38.1	300	2 D64995	probable spermidin
23	45	38.1	539	2 T02508	hypothetical prote
24	45	38.1	542	2 D45557	resB protein - Bac
25	45	38.1	1071	2 F81979	carbamoyl-phosphat
26	45	38.1	1071	2 D81035	carbamoyl-phosphat
27	45	38.1	15281	2 S41309	cyclosporin synthet
28	44.5	37.7	522	2 T43417	response regulator
29	44	37.3	285	2 H70781	hypothetical prote

30	44	37.3	361	2 S57895	hypothetical prote
31	44	37.3	437	2 C86823	GTP-binding protei
32	44	37.3	451	1 S49016	protein-tyrosine k
33	44	37.3	579	2 A27650	phosphate response
34	44	37.3	958	2 T44046	hypothetical prote
35	44	37.3	1410	1 A57013	early endosome ant
36	44	37.3	1474	2 B85188	retrotransposon li
37	44	37.3	1706	2 T39305	protein kinase - f
38	44	37.3	2088	2 E71436	hypothetical prote
39	43.5	36.9	164	2 T44557	hypothetical prote
40	43.5	36.9	326	2 T50497	myosin heavy chain
41	43.5	36.9	556	2 B83847	cytochrome c bioge
42	43	36.4	267	1 DCVKOP	orotidine-5'-phosp
43	43	36.4	286	2 JC2063	gentamicin resista
44	43	36.4	286	2 JE0138	gentamicin-N-acety
45	43	36.4	286	2 S09651	aminoglycoside N3

ALIGNMENTS

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N: Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C: Species: Homo sapiens (man)

C: Date: 30-Jun-1987 #sequence revision 27-Oct-1995 #text change 18-Jun-1999
C: Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140

R: Kornreich, R.; Desnick, R.J.; Bishop, D.F.

Nucleic Acids Res. 17, 3301-3302, 1989

A: Title: Nucleotide sequence of the human alpha-galactosidase A gene.

A: Reference number: S04081; MUID: 89263745; PMID: 2542896

A: Accession: S04081

A: Status: translation not shown

A: Molecule type: DNA

A: Residues: 1-429 <KOI>

A: Cross-references: EMBL: X14448; NID: g31755; PIDN: CAA32617.1; PID: g31756

R: Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.

Gene 58, 177-188, 1987

A: Title: A genomic clone containing the promoter for the gene encoding the human lyso
A: Reference number: A29608; MUID: 88112869; PMID: 2892762

A: Accession: A29608

A: Molecule type: DNA

A: Residues: 1-64 <OUI>

A: Cross-references: GB: M18242; NID: g182944; PIDN: AAA52514.1; PID: g553299

R: Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A: Title: Structural organization of the human alpha-galactosidase A gene: further evi
A: Reference number: A30214; MUID: 88234528; PMID: 2836863

A: Accession: A30214

A: Molecule type: DNA

A: Residues: 1-64 <BIS>

A: Cross-references: EMBL: M20317; EMBL: J03249

R: Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A: Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activit
A: Reference number: S14879; MUID: 90092560; PMID: 2152885

A: Accession: S14879

A: Molecule type: mRNA

A: Residues: 1, 'K', 3-39, 'S', 41-429 <KOI>

A: Cross-references: EMBL: X16889

A: Experimental source: Fabry's disease patient

R: Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick,
Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A: Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding th
A: Reference number: A00896; MUID: 86259694; PMID: 3014515

A: Accession: A00896

A: Molecule type: mRNA

A: Residues: 27-429 <BI>

A: Cross-references: GB: M13571; NID: g178245; PIDN: AAA51676.1; PID: g178246

A: Experimental source: lung

A: Accession: B00896

A: Molecule type: protein

A:Residues: 32-55,'S',57-58,'R',60-65,'S',67-68;228-232,'N',234-237,'A',298-326;'L',334
 R:Suji, S.; Martin, B.M.; Kaslow, D.C.; Migeon, B.R.; Choudary, P.V.; Stubblefield, B.K.
 Eur. J. Biochem. 165, 275-280, 1987
 A:Title: Signal sequence and DNA-mediated expression of human lysosomal alpha-galactosidase
 A:Reference number: I37140; MUID:87246603; PMID:3036505
 A:Accession: I37140
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-429 <RES>
 A:Cross-references: EMBL:X05790; NID:g28535; PIDN:CAA29232.1; PID:g757912
 C:Genetics:
 A:Gene: GDB:GLA
 A:Cross-references: GDB:I19272; OMIM:301500
 A:Map position: Xq21.3-Xq22
 A:Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
 C:Superfamily: alpha-galactosidase
 C:Keywords: Fabry disease; glycolipid metabolism; glycoprotein; glycosidase; hydrolase;
 F.1-31/Domain: signal sequence #status predicted <SIG>
 F.32-429/Product: alpha-galactosidase A #status predicted <MAT>
 F.139,192,215,408/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 75.4%; Score 89; DB 1; Length 429;
 Best Local Similarity 100.0%; Pred. No. 4.7e-06;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHNPTGTVLLQL 18
 |||||
 DB 400 TSRLRSHNPTGTVLLQL 417

RESULT 2
 G95299
 probable adenylate cyclase [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: G95299
 R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 ; Kalman, S.; Keating, D.H.; Palm, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: G95299
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-809 <KUR>
 A:Cross-references: GB:AE006469; PIDN:AAK64961.1; PID:g14523385; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplasmid pSymA
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 688-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma0579
 A:Genome: plasmid

Query Match 46.6%; Score 55; DB 2; Length 809;
 Best Local Similarity 44.4%; Pred. No. 2.3;
 Matches 12; Conservative 5; Mismatches 2; Indels 8; Gaps 1;

QY 3 RLRSHPNTGTVL-----LQSEK 21
 |||||
 DB 486 RLEAHRNPTGLLQQRGPDRIQISER 512

RESULT 3
 D64885
 superinfection exclusion protein B - Escherichia coli (strain K-12)
 C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: D64885; I55010; T09179; S34326
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64885
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-203 <BLAT>
 A:Cross-references: GB:AE000233; GB:U00096; NID:g1787613; PIDN:AAC74435.1; PID:g17876
 A:Experimental source: strain K-12, substrain Mc1655
 R:Faubladier, M.; Bouche, J.P.
 J. Bacteriol. 176, 1150-1156, 1994
 A:Title: Division inhibition gene *difC* of Escherichia coli reveals a widespread group
 A:Reference number: I55010; MUID:94148774; PMID:7508908
 A:Accession: I55010
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 42-116 <RES>
 A:Cross-references: EMBL:Z23096; NID:g312455; PIDN:CAA80642.1; PID:g312456
 A:Experimental source: strain B
 R:Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai,
 .; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Sai
 moto, Y.; Horiuchi, T.
 DNA Res. 3, 363-377, 1996
 A:Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to t
 A:Reference number: Z16603; MUID:97251357; PMID:9097039
 A:Accession: T09179
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 42-203 <AIB>
 A:Cross-references: EMBL:D90774; NID:g1742217; PIDN:BAA14956.1; PID:g1742223
 C:Genetics:
 A:Gene: sieB

Query Match 45.8%; Score 54; DB 2; Length 203;
 Best Local Similarity 51.9%; Pred. No. 0.7;
 Matches 14; Conservative 3; Mismatches 4; Indels 6; Gaps 1;

QY 2 SRLRSHN-----PTGTVLLQLSEKD 22
 |||||
 DB 109 SRLRAHIQKHYSLLPEQVRLLRLSEKE 135.

RESULT 4
 A90871
 phage superinfection exclusion protein [imported] - Escherichia coli (strain O157:H7,
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: A90871
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: A90871
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-203 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA035360.1; PID:g13361402; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 050952
 C:Genetics:
 A:Gene: ECs1937

Query Match 45.8%; Score 54; DB 2; Length 203;
 Best Local Similarity 51.9%; Pred. No. 0.7;
 Matches 14; Conservative 3; Mismatches 4; Indels 6; Gaps 1;

QY 2 SRLRSHN-----PTGTVLLQLSEKD 22
 |||||
 DB 109 SRLRAHIQKHYSLLPEQVRLLRLSEKE 135

RESULT 5

superinfection exclusion protein sieB [similarity] - Escherichia coli (strain O157:H7, S
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
 C:Accession: H85747
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H85747
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-203 <STO>
 A:Cross-references: GB:AE005174; NID:gl2515396; PIDN:AAG56444.1; GSPDB:GN00145; UWGP:224
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: sieB

Query Match 45.8%; Score 54; DB 2; Length 203;

Best Local Similarity 51.9%; Pred. No. 0.7;
 Matches 14; Conservative 3; Mismatches 4; Indels 6; Gaps 1;

QY 2 SRLRSHIN-----PTGTVLLQLSEKD 22
 |||||:| | |||:|||||
 Db 109 SRLRAHIQKHYSLLPEQRVLLRLSEKE 135

RESULT 6

posterior-group protein tudor - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Sep-1999
 C:Accession: A41519; SI9019
 R:Golubeski, G.S.; Bardsley, A.; Tax, F.; Boswell, R.E.
 Genes Dev. 5, 2060-2070, 1991
 A:Title: tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel prote
 A:Reference number: A41519; MUID:92038995; PMID:1936993
 A:Accession: A41519
 A:Molecule type: mRNA
 A:Residues: 1-2515 <GOL>
 A:Cross-references: GB:X62420; NID:g8753; PIDN:CAA44286.1; PID:g8754
 C:Genetics:
 A:Gene: tud
 A:Cross-references: FlyBase:FBgn0003891
 A:Superfamily: posterior-group protein tudor

Query Match 44.1%; Score 52; DB 2; Length 2515;

Best Local Similarity 41.7%; Pred. No. 24;
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQLSEKDEL 24
 | : : | : | |||||
 Db 1981 TKAIITHVENTSRIVLPQSEKDSL 2004

RESULT 7

rapid lysis III protein - phase T4
 N:Alternate names: rIII
 C:Species: phase T4
 C:Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 15-Oct-1999
 C:Accession: JT0765; J00290; S26171; JQ0526
 R:Raudoniklene, A.; Nivinskas, R.
 Gene 134, 135-136, 1993
 A:Title: The sequences of gene rIII of bacteriophage T4 and its mutants.
 A:Reference number: JT0765; MUID:94063508; PMID:8244025
 A:Accession: JT0765
 A:Molecule type: DNA
 A:Residues: 1-82 <RAU>

R:Raudoniklene, A.; Nivinskas, R.
 Nucleic Acids Res. 18, 4280, 1990
 A:Title: Nucleotide sequence of bacteriophage T4 gene 31 region.
 A:Reference number: J00290; MUID:90332452; PMID:2377483
 A:Accession: J00290
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-82 <RA2>

R:Raudoniklene, A.; Nivinskas, R.
 A:Cross-references: EMBL:M37882; NID:g215873; PIDN:AAA32507.1; PID:g215877
 Gene 114, 85-90, 1992
 A:Title: Gene rIII is the nearest downstream neighbour of bacteriophage T4 gene 31.
 A:Reference number: S26167; MUID:92267389; PMID:1587487
 A:Accession: S26171
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-82 <RA3>

A:Cross-references: EMBL:X54536; NID:gl5789; PIDN:CAA38406.1; PID:gl5793
 R:Prillov, A.G.; Mesyanzhinov, V.V.; Aebi, U.; Kellenberger, E.
 Nucleic Acids Res. 18, 3635, 1990
 A:Title: Cloning and sequencing of bacteriophage T4 genes between map positions 128.3
 A:Reference number: JQ0524; MUID:90301484; PMID:2362813
 A:Accession: JQ0526
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-63, 'NLWLLHE', '72', 'QFVTLW' <PRI>
 A:Cross-references: EMBL:X17657; NID:gl5204; PID:gl5208

Query Match 42.8%; Score 50.5; DB 2; Length 82;
 Best Local Similarity 52.2%; Pred. No. 0.91;
 Matches 12; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

QY 3 RLRSINPTGTVL-LQLSEKDEL 24
 | : | | | | : | : | | | | |
 Db 38 RYFKHLNPAQALAAELQEKDEL 50

RESULT 8

JC4522
 alpha-galactosidase (EC 3.2.1.22) A precursor - mouse
 N:Alternate names: alpha-D-galactoside galactohydrolase
 C:Species: Mus musculus (house mouse)
 C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 22-Jun-1999
 C:Accession: JC4522
 R:Ohshima, T.; Murray, G.J.; Nagle, J.W.; Quirk, J.M.; Kraus, M.H.; Barton, N.W.; Bra
 Gene 166, 277-280, 1995
 A:Title: Structural organization and expression of the mouse gene encoding alpha-gala
 A:Reference number: JC4522; MUID:96125203; PMID:8543175
 A:Accession: JC4522
 A:Molecule type: mRNA
 A:Residues: 1-419 <OHS>

A:Cross-references: GB:U34071; NID:gl141787; PIDN:AAA96749.1; PID:gl141788
 A:Experimental source: kidney, C57BL
 C:Comment: This enzyme is a lysosomal enzyme that hydrolyses the alpha-D-galactosyl r
 C:Genetics:
 A:Gene: alpha Gala
 A:Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
 A:Superfamily: alpha-galactosidase
 C:Keywords: glycoprotein; glycosidase; hydrolase; lysosome
 F:1-31/Domain: signal sequence #status predicted <Sig>
 F:32-419/Product: alpha-galactosidase A #status predicted <MAT>
 F:139,192,215,408/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 42.4%; Score 50; DB 2; Length 419;
 Best Local Similarity 50.0%; Pred. No. 6.7;
 Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
 | : : | : | ||||| : |
 Db 400 TLTLKTRVNSGTVLFR 417

RESULT 9

A86372

53.0K hypothetical protein F508_33 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
 C:Accession: A86372
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: A86372
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-475 <STO>
 A:Cross-references: GB:AE005172; NID:g4056460; PIDN:AA098033.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 42.4%; Score 50; DB 2; Length 475;

Best Local Similarity 56.5%; Pred. No. 7.7;
 Matches 13; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 4 LRSHINPT--CTVLLQLSEKDEL 24

DB 27 LRRRINPTVSSVHLSLRKDEL 49

RESULT 10

I40169

carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) - Bacillus caldolyticus
 C:Species: Bacillus caldolyticus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000

C:Accession: I40169; S34321

R:Ghim, S.Y.; Neuhaud, J.

J. Bacteriol. 176, 3698-3707, 1994

A:Title: The pyrimidine biosynthesis operon of the thermophile Bacillus caldolyticus and its regulation.

A:Reference number: I40166; MUID:94266723; PMID:8206848

A:Accession: I40169

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1065 <RES>

A:Cross-references: EMBL:X73308; NID:g312439; PIDN:CAA51739.1; PID:g312443

C:Genetics:

A:Gene: pyrB

C:Function:

A:Pathway: glutamate metabolism; pyrimidine nucleotide biosynthesis
 C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis
 F:9-1047/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homodimer
 F:9-467/Domain: biotin carboxylase homology <BC1>
 F:553-985/Domain: biotin carboxylase homology <BC2>

Query Match 42.4%; Score 50; DB 1; Length 1065;

Best Local Similarity 52.9%; Pred. No. 19;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 7 HINPTGTVLLQLSEKDE 23

DB 931 HIQPHGAVLLTVADKDK 947

RESULT 11

E91088

DNA helicase RecB [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: E91088

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Gasawara, N.; Vasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome organization of the verotoxin genes.

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: E91088

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1180 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA037100.1; PID:g13363149; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 41.5%; Score 49; DB 2; Length 1180;

Best Local Similarity 43.5%; Pred. No. 31;

Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQLSEKDE 23

DB 995 TAVLQAPLNETGVSLQSLSDRDK 1017

RESULT 12

G85933

DNA helicase RecB [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001

C:Accession: G85933

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85933

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1180 <STO>

A:Cross-references: GB:AE005174; NID:g12517302; PIDN:AAG57931.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: recB

C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 41.5%; Score 49; DB 2; Length 1180;

Best Local Similarity 43.5%; Pred. No. 31;

Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQLSEKDE 23

DB 995 TAVLQAPLNETGVSLQSLSDRDK 1017

RESULT 13

D81028

Maf/YceF/YhdE family protein NMB1909 [imported] - Neisseria meningitidis (strain MC58)

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 27-Nov-2001

C:Accession: D81028

R:Tattalin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; H. H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M. Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: D81028

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-196 <DET>

A:Cross-references: GB:AE002540; GB:AE002098; NID:g7227159; PIDN:AAF42239.1; PID:g722

A:Experimental source: serogroup B, strain MC58

Search completed: July 29, 2003, 16:18:57
Job time : 4.68456 secs

C:Genetics:

A:Gene: NM1909

C:Superfamily: septum formation protein maf

Query Match 40.7%; Score 48; DB 2; Length 196;
Best Local Similarity 50.0%; Pred. No. 6;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 1 TSLRLSHINPTGTVLLQLSE 20

I : I : I : I : I : I : I

Db 116 TGRMRHDKTVVVMRQLDE 135

RESULT 14

AG1793

para-aminobenzoate synthase component I homolog lin2893 [imported] - Listeria innocua (S

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG1793

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourham, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tisseret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1793

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-568 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC98119.1; PID:g16415428; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2893

Query Match

39.8%; Score 47; DB 2; Length 568;

Best Local Similarity 50.0%; Pred. No. 28;

Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TSLRLSHINPTGTVLLQLSE 20

I : I : I : I : I : I : I

Db 423 TYKRLFLHPNGTHLELAE 442

RESULT 15

C69264

conserved hypothetical protein AF0115 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: C69264

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: C69264

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-257 <KLE>

A:Cross-references: GB:AF001098; GB:AE000782; NID:g2689421; PIDN:AAB91113.1; PID:g265053

C:Superfamily: hypothetical protein YLR351c

Query Match

39.0%; Score 46; DB 2; Length 257;

Best Local Similarity 52.9%; Pred. No. 17;

Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 8 INPTGTVLLQLSEKDEL 24

I : I : I : I : I : I : I

Db 213 IDPMGTVLIEASEYQEV 229

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:17:20 ; Search time 3.65101 Seconds
(without alignments)
780.671 Million cell updates/sec

Title: US-09-993-059-30

Perfect score: 118

Sequence: 1 TSRLSRHINPTGVLLQLSEKDEL 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	100.0	24	15	US-10-103-327-30
2	118	100.0	423	15	US-10-103-327-18
3	106	89.8	427	15	US-10-103-327-14
4	104	88.1	32	15	US-10-103-327-26
5	104	88.1	431	15	US-10-103-327-10
6	102	86.4	36	15	US-10-103-327-24
7	102	86.4	435	15	US-10-103-327-6
8	95.5	80.9	29	15	US-10-103-327-28
9	89	75.4	18	15	US-10-103-327-29
10	89	75.4	22	15	US-10-103-327-27
11	89	75.4	26	15	US-10-103-327-25
12	89	75.4	30	15	US-10-103-327-23
13	89	75.4	398	15	US-10-165-060-4
14	89	75.4	398	15	US-10-165-968-4
15	89	75.4	417	15	US-10-103-327-16

16	75.4	421	15	US-10-103-327-12	Sequence 12, Appl
17	75.4	424	15	US-10-103-327-8	Sequence 8, Appl
18	75.4	428	15	US-10-103-327-4	Sequence 4, Appl
19	45.8	203	9	US-09-912-020-338	Sequence 338, App
20	38.1	61	15	US-10-156-761-12546	Sequence 12546, A
21	36.4	82	15	US-10-242-943-26	Sequence 26, Appl
22	36.4	299	15	US-10-023-282-265	Sequence 265, App
23	36.4	711	9	US-09-966-561-2	Sequence 2, Appl
24	35.6	72	10	US-09-945-182-11	Sequence 11, Appl
25	35.6	72	10	US-09-945-182-13	Sequence 13, Appl
26	35.6	240	10	US-09-945-182-30	Sequence 30, Appl
27	35.6	263	10	US-09-945-182-32	Sequence 32, Appl
28	35.6	294	10	US-09-945-182-2	Sequence 2, Appl
29	35.6	321	10	US-09-945-182-26	Sequence 26, Appl
30	35.6	388	10	US-09-945-182-34	Sequence 34, Appl
31	35.6	411	10	US-09-945-182-28	Sequence 28, Appl
32	35.6	436	9	US-09-730-772-14	Sequence 14, Appl
33	35.6	436	9	US-09-735-849-14	Sequence 14, Appl
34	35.6	450	15	US-10-188-246-12	Sequence 12, Appl
35	35.6	600	15	US-10-156-761-8406	Sequence 8406, Ap
36	35.6	665	15	US-10-121-393-2	Sequence 2, Appl
37	35.6	2092	15	US-10-128-714-3061	Sequence 3061, Ap
38	35.6	2092	15	US-10-128-714-8061	Sequence 8061, Ap
39	41.5	161	15	US-10-156-761-12336	Sequence 12336, A
40	34.7	57	15	US-10-156-761-10230	Sequence 10230, A
41	34.7	72	10	US-09-945-182-15	Sequence 15, Appl
42	34.7	98	15	US-10-222-334-5	Sequence 5, Appl
43	34.7	119	8	US-08-945-459A-1	Sequence 1, Appl
44	34.7	119	10	US-09-880-708-13	Sequence 13, Appl
45	34.7	119	10	US-09-068-253-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-103-327-30
; Sequence 30, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-30

Query Match 100.0%; Score 118; DB 15; Length 24;
Best Local Similarity 100.0%; Pred. No. 8e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 TSRLSRHINPTGVLLQLSEKDEL 24
|||||
Db 1 TSRLSRHINPTGVLLQLSEKDEL 24

RESULT 2
US-10-103-327-18
; Sequence 18, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match 100.0%; Score 118; DB 15; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQSEKDEL 24
Db 400 TSLRSHINPTGTVLLQSEKDEL 423

RESULT 3

US-10-103-327-14
; Sequence 14, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match 89.8%; Score 106; DB 15; Length 427;
Best Local Similarity 85.7%; Pred. No. 1.7e-08;
Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 TSLRSHINPTGTVLLQ-----SEKDEL 24
Db 400 TSLRSHINPTGTVLLQLENTMQMSLSEKDEL 427

RESULT 4

US-10-103-327-26
; Sequence 26, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-26

Query Match 88.1%; Score 104; DB 15; Length 32;
Best Local Similarity 75.0%; Pred. No. 1.8e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 TSLRSHINPTGTVLLQ-----LSEKDEL 24
Db 1 TSLRSHINPTGTVLLQLENTMQMSLSEKDEL 32

RESULT 5

US-10-103-327-10
; Sequence 10, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-10

Query Match 88.1%; Score 104; DB 15; Length 431;
Best Local Similarity 75.0%; Pred. No. 3.6e-08;
Matches 24; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 TSLRSHINPTGTVLLQ-----LSEKDEL 24
Db 400 TSLRSHINPTGTVLLQLENTMQMSLSEKDEL 431

RESULT 6

US-10-103-327-24
; Sequence 24, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-24

Query Match 86.4%; Score 102; DB 15; Length 36;

Best Local Similarity 66.7%; Pred. No. 4.2e-09;
Matches 24; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 TSRLRSHNPCTGVLLQ-----LSEKDEL 24
|||||
DB 1 TSRLRSHNPCTGVLLQLENTMQMSLKDLLSEKDEL 36
|||||

RESULT 7

US-10-103-327-6

; Sequence 6, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 435

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-103-327-6

Query Match

Best Local Similarity 86.4%; Score 102; DB 15; Length 435;

Matches 24; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 TSRLRSHNPCTGVLLQ-----LSEKDEL 24
|||||
DB 400 TSRLRSHNPCTGVLLQLENTMQMSLKDLLSEKDEL 435
|||||

RESULT 8

US-10-103-327-28

; Sequence 28, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 28

; LENGTH: 29

; TYPE: PRT

; ORGANISM: Tobacco mosaic virus

US-10-103-327-28

Query Match

Best Local Similarity 80.9%; Score 95.5; DB 15; Length 29;

Matches 24; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 1 TSRLRSHNPCTGVLLQ-----SEKDEL 24
|||||
DB 1 TSRLRSHNPCTGVLLQLENTMSEKDEL 29
|||||

RESULT 9

US-10-103-327-29

; Sequence 29, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 29

; LENGTH: 18

; TYPE: PRT

; ORGANISM: Tobacco mosaic virus

US-10-103-327-29

Query Match

Best Local Similarity 75.4%; Score 89; DB 15; Length 18;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHNPCTGVLLQL 18
|||||
DB 1 TSRLRSHNPCTGVLLQL 18
|||||

RESULT 10

US-10-103-327-27

; Sequence 27, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 27

; LENGTH: 22

; TYPE: PRT

; ORGANISM: Tobacco mosaic virus

US-10-103-327-27

Query Match

Best Local Similarity 75.4%; Score 89; DB 15; Length 22;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHNPCTGVLLQL 18
|||||
DB 1 TSRLRSHNPCTGVLLQL 18
|||||

RESULT 11

US-10-103-327-25

; Sequence 25, Application US/10103327

; Publication No. US20030106095A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-25

Query Match 75.4%; Score 89; DB 15; Length 26;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18

RESULT 12

US-10-103-327-23
; Sequence 23, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-23

Query Match 75.4%; Score 89; DB 15; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.8e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18

RESULT 13

US-10-165-060-4
; Sequence 4, Application US/10165060
; Publication No. US20030077806A1
; GENERAL INFORMATION:

; APPLICANT: Seiden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,060
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041

; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-165-060-4

Query Match 75.4%; Score 89; DB 15; Length 398;
Best Local Similarity 100.0%; Pred. No. 7.4e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQL 18
|||||
Db 369 TSLRSHINPTGTVLLQL 386

RESULT 14

US-10-165-968-4
; Sequence 4, Application US/10165968
; Publication No. US20030113894A1
; GENERAL INFORMATION:

; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,968
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-165-968-4

Query Match 75.4%; Score 89; DB 15; Length 398;
Best Local Similarity 100.0%; Pred. No. 7.4e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLLQL 18
|||||
Db 369 TSLRSHINPTGTVLLQL 386

RESULT 15

US-10-103-327-16
; Sequence 16, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

```
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match      75.4%; Score 89; DB 15; Length 417;
Best Local Similarity 100.0%; Pred. No. 7.8e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TSRLRSHINPTGTVLLQL 18
        |||
Db      400 TSRLRSHINPTGTVLLQL 417
```

Search completed: July 29, 2003, 16:29:36
Job time : 5.15101 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 16:09:29 ; Search time 2.20134 seconds
(without alignments)
461.292 Million cell updates/sec

Title: US-09-993-059-30
Perfect score: 118
Sequence: 1 TSRLRSHINPTGVLLQLSEKDEL 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	75.4	92	1 US-07-602-824A-9	Sequence 9, Appli
2	89	75.4	92	1 US-07-983-451-9	Sequence 9, Appli
3	89	75.4	92	1 US-08-261-577-12	Sequence 12, Appl
4	89	75.4	386	3 US-09-176-666-11	Sequence 11, Appl
5	89	75.4	387	3 US-09-176-666-10	Sequence 10, Appl
6	89	75.4	388	3 US-09-176-666-9	Sequence 9, Appli
7	89	75.4	389	3 US-09-176-666-8	Sequence 8, Appli
8	89	75.4	390	3 US-09-176-666-7	Sequence 7, Appli
9	89	75.4	391	3 US-09-176-666-6	Sequence 6, Appli
10	89	75.4	392	3 US-09-176-666-5	Sequence 5, Appli
11	89	75.4	393	3 US-09-176-666-4	Sequence 4, Appli
12	89	75.4	394	3 US-09-176-666-3	Sequence 3, Appli
13	89	75.4	396	3 US-09-176-666-2	Sequence 2, Appli
14	89	75.4	398	3 US-08-928-881-26	Sequence 26, Appl
15	89	75.4	398	3 US-09-176-666-1	Sequence 1, Appli
16	89	75.4	398	4 US-09-543-921-26	Sequence 26, Appl
17	89	75.4	398	4 US-09-266-014-4	Sequence 4, Appli
18	89	75.4	398	4 US-09-491-759-26	Sequence 26, Appl
19	89	75.4	428	6 5179023-4	Patent No. 5179023
20	89	75.4	429	1 US-07-602-824A-2	Sequence 2, Appli
21	89	75.4	429	1 US-07-602-608-3	Sequence 3, Appli
22	89	75.4	429	1 US-07-983-451-2	Sequence 2, Appli
23	89	75.4	429	1 US-08-261-578-3	Sequence 3, Appli
24	89	75.4	429	1 US-08-261-577-7	Sequence 7, Appli
25	89	75.4	429	1 US-08-261-577-9	Sequence 9, Appli
26	89	75.4	429	3 US-09-070-356-4	Sequence 4, Appli
27	68	57.6	381	3 US-09-176-666-12	Sequence 12, Appl

28	45	38.1	15281	2	US-08-471-119A-2	Sequence 2, Appli
29	43.5	36.9	249	4	US-09-252-991A-27447	Sequence 27447, A
30	43	36.4	82	4	US-09-167-206-26	Sequence 26, Appl
31	43	36.4	270	2	US-08-859-201-4	Sequence 4, Appli
32	43	36.4	299	4	US-09-205-258-265	Sequence 265, App
33	43	36.4	314	2	US-08-859-201-8	Sequence 8, Appli
34	43	36.4	482	4	US-09-107-532A-4512	Sequence 2, Appli
35	43	36.4	714	2	US-08-859-201-2	Sequence 2, Appli
36	43	36.4	763	4	US-09-107-532A-4831	Sequence 4831, Ap
37	43	36.4	893	1	US-07-977-434-4	Sequence 4, Appli
38	43	36.4	893	1	US-08-458-819-4	Sequence 4, Appli
39	43	36.4	893	3	US-09-103-697-10	Sequence 10, Appl
40	43	36.4	893	5	PCT-US91-07035-4	Sequence 4, Appli
41	42	35.6	72	1	US-08-362-670B-11	Sequence 11, Appl
42	42	35.6	72	1	US-08-362-670B-13	Sequence 13, Appl
43	42	35.6	72	3	US-08-333-576C-11	Sequence 11, Appl
44	42	35.6	72	3	US-08-333-576C-13	Sequence 13, Appl
45	42	35.6	72	3	US-08-808-324-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-07-602-824A-9
; Sequence 9, Application US/07602824A
; Patent No. 5356804
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; NUMBER OF INVENTIONS: ACTIVE alpha-GALACTOSIDASE A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,824A
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-9

Query Match 75.4%; Score 89; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGVLLQL 18
DB 3 TSRLRSHINPTGVLLQL 20

RESULT 2
US-07-983-451-9
; Sequence 9, Application US/07983451
; Patent No. 5401650
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/983,451
; FILING DATE: 30-NOV-1992
; CLASSIFICATION: I 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 07/983,451
; REFERENCE/DOCKET NUMBER: 6923-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-983-451-9
Query Match 75.4%; Score 89; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 3 TSRLRSHINPTGTVLLQL 20
RESULT 3
US-08-261-577-12
; Sequence 12, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-261-577-12
Query Match 75.4%; Score 89; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 3 TSRLRSHINPTGTVLLQL 20
RESULT 4
US-09-176-666-11
; Sequence 11, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-11
Query Match 75.4%; Score 89; DB 3; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386
RESULT 5
US-09-176-666-10
; Sequence 10, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666

; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-10

Query Match 75.4%; Score 89; DB 3; Length 387;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 6
US-09-176-666-9
; Sequence 9, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-9

Query Match 75.4%; Score 89; DB 3; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 7
US-09-176-666-8
; Sequence 8, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-8

Query Match 75.4%; Score 89; DB 3; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 8
US-09-176-666-7
; Sequence 7, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-7

Query Match 75.4%; Score 89; DB 3; Length 390;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 9
US-09-176-666-6
; Sequence 6, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-6

Query Match 75.4%; Score 89; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 10
US-09-176-666-5
; Sequence 5, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro

; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-5

Query Match 75.4%; Score 89; DB 3; Length 392;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386
|||||

RESULT 11
US-09-176-666-4
; Sequence 4, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-4

Query Match 75.4%; Score 89; DB 3; Length 393;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386
|||||

RESULT 12
US-09-176-666-3
; Sequence 3, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-176-666-3

Query Match 75.4%; Score 89; DB 3; Length 394;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386
|||||

RESULT 13
US-09-176-666-2
; Sequence 2, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-2

Query Match 75.4%; Score 89; DB 3; Length 396;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
Db 369 TSRLRSHINPTGTVLLQL 386
|||||

RESULT 14
US-08-928-881-26
; Sequence 26, Application US/08928881
; Patent No. 6083725
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,881
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906

; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-928-881-26

Query Match 75.4%; Score 89; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
|||||
Db 369 TSRLRSHINPTGTVLLQL 386

RESULT 15

US-09-176-666-1
; Sequence 1, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-176-666-1

Query Match 75.4%; Score 89; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
|||||
Db 369 TSRLRSHINPTGTVLLQL 386

Search completed: July 29, 2003, 16:17:58
Job time : 3.20134 secs

THIS PAGE BLANK (uspto)